



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Sven Enerbäck et al.  
Serial No. : 09/963,285  
Filed : September 26, 2001  
Title : PROMOTER SEQUENCES

Art Unit : 1645  
Examiner :

Commissioner for Patents  
Washington, D.C. 20231

TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT UNDER 35 USC §119

In accordance with the provisions of 35 U.S.C. §119, applicants hereby claim priority of Swedish Patent application No. 0004102-0, filed November 9, 2000. A certified copy of the application is submitted herewith. As the priority application is in the English language, all of the requirements of §119 have been met.

Please apply any charges or credits to Deposit Account No. 06-1050, referencing Attorney Docket No. 13425-042001.

Respectfully submitted,

Date:

January 22, 2002

Jack Brennan

Jack Brennan  
Reg. No. 47,443

Fish & Richardson P.C.  
225 Franklin Street  
Boston, Massachusetts 02110-2804  
Telephone: (617) 542-5070  
Facsimile: (617) 542-8906

20376379.doc

CERTIFICATE OF MAILING BY FIRST CLASS MAIL

I hereby certify under 37 CFR §1.8(a) that this correspondence is being deposited with the United States Postal Service as first class mail with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit

January 22, 2002

Signature

Darlene J. Morin

Typed or Printed Name of Person Signing Certificate

Darlene J. Morin

# PRV

PATENT- OCH REGISTRERINGSVERKET

Patentavdelningen



## Intyg Certificat



Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

*This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.*

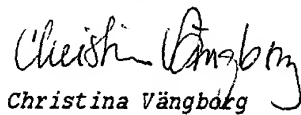
(71) Sökande Biovitrum AB, Stockholm SE  
Applicant (s)

(21) Patentansökningsnummer 0004102-0  
Patent application number

(86) Ingivningsdatum 2000-11-09  
Date of filing

Stockholm, 2001-11-16

För Patent- och registreringsverket  
For the Patent- and Registration Office

  
Christina Vängborg

Avgift  
Fee 170:-

**PATENT- OCH  
REGISTRERINGSVERKET  
SWEDEN**

Postadress/Adress  
Box 5055  
S-102 42 STOCKHOLM

Telefon/Phone  
+46 8 782 25 00  
Vx 08-782 25 00

Telex  
17978  
PATOREG S

Telefax  
+46 8 666 02 66  
08-666 02 66

## PROMOTER SEQUENCES II

### TECHNICAL FIELD

The present invention relates an isolated promoter region of the mammalian transcription  
5 factor *FOXC2*. The invention also relates to screening methods for agents modulating the  
expression of *FOXC2* and thereby being potentially useful for the treatment of medical  
conditions related to obesity. The invention further relates to a previously unknown variant  
of the human *FOXC2* gene, derived via the use of an alternative promoter, which produces  
an additional exon that generates a distinct open reading frame via splicing. The alternative  
10 gene encodes a variant of the *FOXC2* transcription factor, which is lacking a part of the  
DNA-binding domain and consequently has a potential regulatory function.

### BACKGROUND ART

15 More than half of the men and women in the United States, 30 years of age and older, are  
now considered overweight, and nearly one-quarter are clinically obese. This high  
prevalence has led to increases in the medical conditions that often accompany obesity,  
especially non-insulin dependent diabetes mellitus (NIDDM), hypertension, cardiovascular  
disorders, and certain cancers. Obesity results from a chronic imbalance between energy  
20 intake (feeding) and energy expenditure. To better understand the mechanisms that lead to  
obesity and to develop strategies in certain patient populations to control obesity, there is a  
need to develop a better underlying knowledge of the molecular events that regulate the  
differentiation of preadipocytes and stem cells to adipocytes, the major component of  
adipose tissue.

25 The helix-loop-helix (HLH) family of transcriptional regulatory proteins are key players in  
a wide array of developmental processes (for a review, see Massari & Murre (2000) Mol.  
Cell. Biol. 20: 429-440). Over 240 HLH proteins have been identified to date in organisms  
ranging from the yeast *Saccharomyces cerevisiae* to humans. Studies in *Xenopus laevis*,  
30 *Drosophila melanogaster*, and mice have convincingly demonstrated that HLH proteins are  
intimately involved in developmental events such as cellular differentiation, lineage

commitment, and sex determination. In multicellular organisms, HLH factors are required for a multitude of important developmental processes, including neurogenesis, myogenesis, hematopoiesis, and pancreatic development.

- 5 The winged helix / forkhead class of transcription factors is characterized by a 100-amino acid, monomeric DNA-binding domain. X-ray crystallography of the forkhead domain from HNF-3 $\gamma$  has revealed a three-dimensional structure, the "winged helix", in which two loops (wings) are connected on the C-terminal side of the helix-loop-helix (for reviews, see Brennan, R.G. (1993) Cell 74: 773-776; and Lai, E. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90: 10421-10423).
- 10

- The isolation of the mouse mesenchyme forkhead-1 (MFH-1) and the corresponding human (*FKHL14*) chromosomal genes is disclosed by Miura, N. et al. (1993) FEBS letters 326: 171-176; and (1997) Genomics 41: 489-492. The nucleotide sequences of the mouse
- 15 MFH-1 gene and the human *FKHL14* gene have been deposited with the EMBL/GenBank Data Libraries under accession Nos. Y08222 (SEQ ID NO: 5) and Y08223 (SEQ ID NO: 8), respectively. A corresponding gene has been identified in *Gallus gallus* (GenBank accession numbers U37273 and U95823).

- 20 The International Patent Application WO 98/54216 discloses a gene encoding a Forkhead-Related Activator (FREAC)-11 (also known as S12), which is identical with the polypeptide encoded by the human *FKHL14* gene disclosed by Miura, *supra*. This transcription factor is expressed in adipose tissue and involved in lipid metabolism and adipocyte differentiation (cf. Swedish patent application No. 0000531-4, filed February 18,
- 25 2000).

- The nomenclature for the winged helix / forkhead transcription factors has been standardized and Fox (Forkhead Box) has been adopted as the unified symbol (Kaestner et al. (2000) Genes & Development 14: 142-146; see also <http://www.biology.pomona.edu/fox>). It has been agreed that the genes previously designated MFH-1 and *FKHL14* (as well as FREAC-11 and S12) should be designated *FOXC2*.
- 30

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the general structure of the human *FOXC2* gene.

- 5 Figure 2 illustrates the results from phylogenetic footprinting experiments. Shown is the fraction conserved (1.0 = 100%) between mouse *FoxC2* and human *FOXC2* sequences in the alignment generated with Clustal. Solid (bold) line indicates the fraction of the human sequence which is identical to the mouse within a 200 bp "window" over the human sequence in the alignment. The weak (dotted) line is set to -0.05 when the sliding window contains human exon sequence and to -0.1 when the window is entirely composed of exon sequence. Regions containing local maxima or exceeding a conservation fraction of 0.7 are likely to be functional and are classified as "predicted regulatory regions".
- 10

- Figure 3 illustrates the predicted "enhancer" region in the human *FOXC2* gene. Underlined sequences indicate likely transcription factor binding sites. Boxed sequence indicates exon sequence.
- 15

*Splice* = sequence predicted as splice site in the alternatively spliced gene;

- E-box-like* = sequence resembling the "E-box" motif CANNTG known as a target for DNA binding proteins containing a helix-loop-helix domain (often associated with the activation of cell-type specific gene transcription during tissue differentiation; see Massari & Murre (2000) Mol. Cell. Biol. 20: 429-440)
- 20

*Forkhead-like* = sequence resembling binding site for the winged helix / forkhead class of transcription factors;

- Ets-like* = sequence resembling consensus binding site for ETS-domain transcription factor family (see Sharrocks et al. (1997) Int. J. Biochem. Cell Biol. 29, 1371-1387).
- 25

Figure 4 illustrates the predicted "promoter" region in the human *FOXC2* gene. Underlined sequence indicates exon sequences. Boxed sequences indicate conserved block (potential transcription factor binding sites).

30

## DESCRIPTION OF THE INVENTION

According to the present invention, the partially known sequence (SEQ ID NO: 8) of human *FOXC2* gene has been extended. In the previously unknown region of the gene, differentially conserved regions, consistent with regulatory function, have been identified. Further, an alternative transcript has been identified, which includes the use of at least two exons. The putative regulatory enhancer is immediately adjacent to the newly discovered alternative exon, suggesting that it may play a role in the alternative selection of transcript classes.

10

Modulation of the *FOXC2* regulation is expected to have therapeutic value in type II diabetes; obesity, hypercholesterolemia, and other cardiovascular diseases or dyslipidemias.

15 Consequently, in a first aspect this invention provides a human *FOXC2* promoter region comprising a sequence selected from:

- (a) the nucleotide sequence set forth as positions 1250 to 2235, such as positions 1250 to 1749 or positions 1692 to 1703, in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* promoter activity;
- 20 (b) the complementary strand of (a); and
- (c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).

Another aspect of the invention is a recombinant construct comprising the human *FOXC2* promoter region as defined above. In the said recombinant construct, the human *FOXC2* promoter region can be operably linked to a gene encoding a detectable product, such as the human *FOXC2* gene, or a reporter gene. The term "operably linked" as used herein means functionally fusing a promoter with a structural gene in the proper frame to express the structural gene under control of the promoter. As used herein, the term "reporter gene" means a gene encoding a gene product that can be identified using simple, inexpensive methods or reagents and that can be operably linked to the human *FOXC2* promoter region or an active fragment thereof. Reporter genes such as, for example, a luciferase,  $\beta$ -galactosidase, alkaline phosphatase, or green fluorescent protein reporter gene, can be used

25

30

to determine transcriptional activity in screening assays according to the invention (see, for example, Goeddel (ed.), *Methods Enzymol.*, Vol. 185, San Diego: Academic Press, Inc. (1990); see also Sambrook, *supra*).

- 5 The invention also provides a vector comprising the recombinant construct as defined above, as well as a host cell stably transformed with such a vector, or generally with the recombinant construct according to the invention. The term "vector" refers to any carrier of exogenous DNA that is useful for transferring the DNA to a host cell for replication and/or appropriate expression of the exogenous DNA by the host cell.

10

In another aspect, the invention provides a method for identification of an agent regulating *FOXC2* promoter activity, said method comprising the steps: (i) contacting a candidate agent with a human *FOXC2* promoter region as defined above; and (ii) determining whether said candidate agent modulates expression of the *FOXC2* gene, such modulation  
15 being indicative for an agent capable of regulating *FOXC2* promoter activity. As used herein, the term "agent" means a biological or chemical compound such as a simple or complex organic molecule, a peptide, a protein or an oligonucleotide.

- A transfection assay can be a particularly useful screening assay for identifying an  
20 effective agent modulating and/or regulating *FOXC2* promoter activity. In a transfection assay, a nucleic acid containing a gene, e.g. a reporter gene, operably linked to a human *FOXC2* promoter or an active fragment thereof, is transfected into the desired cell type. A test level of reporter gene expression is assayed in the presence of a candidate agent and compared to a control level of expression. An effective agent is identified as an agent that  
25 results in a test level of expression that is different than a control level of reporter gene expression, which is the level of expression determined in the absence of the agent. Methods for transfecting cells and a variety of convenient reporter genes are well known in the art (see, for example, Goeddel (ed.), *Methods Enzymol.*, Vol. 185, San Diego: Academic Press, Inc. (1990); see also Sambrook, *supra*). Consequently, the said method  
30 could e.g. comprising assaying reporter gene expression in a host cell, stably transformed with a recombinant construct comprising the human *FOXC2* promoter, in the presence and absence of a candidate agent, wherein an effect on the test level of expression as compared

to control level of expression is indicative of an agent capable of regulating *FOXC2* promoter activity.

Methods for identification of polypeptides regulating *FOXC2* promoter activity could  
5 include various techniques known in the art, such as the yeast one-hybrid system (see: Li & Herskowitz (1993) Science 262, 1870-1874) to identify proteins binding specific sequences from the *FOXC2* regulatory region, biochemical purification of proteins which bind to the regulatory region, the use of a "southwestern" cloning strategy (see e.g. Hai et al. (1989) Genes & Development 3: 2083-2090) in which a pool of bacteria infected with a  
10 "phage library" are induced to express the encoded protein and probed with radioactive DNA sequences from the *FOXC2* regulatory regions to identify binding proteins.

In a further aspect, the invention provides a human *FOXC2* enhancer region comprising a sequence selected from:

- 15 (a) the nucleotide sequence set forth as positions 216 to 475, such as positions 223 to 231, positions 359 to 375, positions 378 to 402, or positions 403 to 423, in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;  
(b) the complementary strand of (a); and  
(c) nucleotide sequences capable of hybridizing, under stringent hybridization  
20 conditions, to a nucleotide sequence as defined in (a) or (b).

As described above for the human *FOXC2* promoter region, the invention further provides a recombinant construct comprising a human *FOXC2* enhancer region, a vector comprising the said recombinant construct, as well as a host cell stably transformed with said vector or  
25 with said recombinant construct.

Further, the invention provides a method for identification of an agent regulating *FOXC2* enhancer activity, said method comprising the steps: (i) contacting a candidate agent with the human *FOXC2* enhancer region as defined above; and (ii) determining whether said  
30 candidate agent modulates expression of the *FOXC2* gene, such modulation being indicative for an agent capable of regulating *FOXC2* enhancer activity. It will be understood by the skilled person that known steps are available for performing such a method. For instance, a "panel" of constructs which include a variety of mutations and



deletions can be used in order to associate a response with a specific alteration of a single base or subsegment of the regulatory apparatus. A simple panel might include: enhancer plus promoter, promoter only, enhancer plus a "minimal" promoter from a distinct gene. As mentioned above, a transfection assay, using a host cell stably transformed with a suitable recombinant construct, can be a particularly useful screening assay for identifying an effective agent.

In yet a further aspect, the invention provides a method for identification of an agent capable of regulating a mammalian *FOXC2* promoter activity, said method comprising the steps (i) contacting a candidate agent with a murine *FoxC2* promoter nucleotide sequence shown as positions 216 to 2235, such as positions 216 to 475 or positions 1250 to 2235, in SEQ ID NO: 5; and (ii) determining whether said candidate agent modulates expression of a mammalian *FOXC2* gene, such modulation being indicative for an agent capable of regulating mammalian *FOXC2* promoter activity.

15

In another important aspect, the invention provides an isolated nucleic acid molecule selected from:

- (a) nucleic acid molecules comprising a nucleotide sequence as shown in SEQ ID NO: 3;
- (b) nucleic acid molecules comprising a nucleotide sequence capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence complementary the polypeptide coding region of a nucleic acid molecule as defined in (a) and which codes for a variant form of the *FOXC2* transcription factor; and
- (c) nucleic acid molecules comprising a nucleic acid sequence which is degenerate as a result of the genetic code to a nucleotide sequence as defined in (a) or (b) and which codes for a variant form of the *FOXC2* transcription factor.

25

In a preferred form of the invention, the said nucleic acid molecule has a nucleotide sequence identical with SEQ ID NO: 3 of the Sequence Listing. However, the nucleic acid molecule according to the invention is not to be limited strictly to the sequence shown as SEQ ID NO: 3. Rather the invention encompasses nucleic acid molecules carrying modifications like substitutions, small deletions, insertions or inversions, which nevertheless encode proteins having substantially the biochemical activity of the *FOXC2* polypeptide according to the invention. Included in the invention are consequently nucleic

30

acid molecules, the nucleotide sequence of which is at least 90% homologous, preferably at least 95% homologous, with the nucleotide sequence shown as SEQ ID NO: 3 in the Sequence Listing.

- 5 Included in the invention is also a nucleic acid molecule which nucleotide sequence is degenerate, because of the genetic code, to the nucleotide sequence shown as SEQ ID NO: 3. A sequential grouping of three nucleotides, a "codon", codes for one amino acid. Since there are 64 possible codons, but only 20 natural amino acids, most amino acids are coded for by more than one codon. This natural "degeneracy", or "redundancy", of the genetic  
10 code is well known in the art. It will thus be appreciated that the nucleotide sequence shown in the Sequence Listing is only an example within a large but definite group of sequences which will encode the variant FOXC2 polypeptide.

- The invention includes an isolated polypeptide encoded by the nucleic acid as defined  
15 above. In a preferred form, the said polypeptide has an amino acid sequence according to SEQ ID NO: 4 of the Sequence Listing. However, the polypeptide according to the invention is not to be limited strictly to a polypeptide with an amino acid sequence identical with SEQ ID NO: 4 in the Sequence Listing. Rather the invention encompasses polypeptides carrying modifications like substitutions, small deletions, insertions or  
20 inversions, which polypeptides nevertheless have substantially the biological activities of the variant FOXC2 polypeptide.

- A further aspect of the invention is a vector harboring the nucleic acid molecule according to the invention. The said vector can e.g. be a replicable expression vector, which carries  
25 and is capable of mediating the expression of a DNA molecule according to the invention. In the present context the term "replicable" means that the vector is able to replicate in a given type of host cell into which it has been introduced. Examples of vectors are viruses such as bacteriophages, cosmids, plasmids and other recombination vectors. Nucleic acid molecules are inserted into vector genomes by methods well known in the art.

- 30 Included in the invention is also a cultured host cell harboring a vector according to the invention. Such a host cell can be a prokaryotic cell, a unicellular eukaryotic cell or a cell derived from a multicellular organism. The host cell can thus e.g. be a bacterial cell such as

an *E. coli* cell; a cell from yeast such as *Saccharomyces cerevisiae* or *Pichia pastoris*, or a mammalian cell. The methods employed to effect introduction of the vector into the host cell are standard methods well known to a person familiar with recombinant DNA methods.

5

In yet another aspect, the invention includes a method for identifying an agent capable of regulating expression of the nucleic acid molecule as defined above, said method comprising the steps (i) contacting a candidate agent with the said nucleic acid molecule; and (ii) determining whether said candidate agent modulates expression of the said nucleic acid molecule.

10

In another aspect the invention provides an antisense oligonucleotide having a sequence capable of specifically hybridizing to RNA transcribed by the alternatively spliced nucleic acid molecule shown as SEQ ID NO: 3, so as to prevent translation of the said RNA.

15

Antisense nucleic acids (preferably 10 to 20 base-pair oligonucleotides) capable of specifically binding to control sequences for the alternatively spliced *FOXC2* gene are introduced into cells, e.g. by a viral vector or colloidal dispersion system such as a liposome. The antisense nucleic acid binds to the target nucleotide sequence in the cell and prevents transcription and/or translation of the target sequence. Phosphorothioate and methylphosphonate antisense oligonucleotides are specifically contemplated for therapeutic use by the invention. Suppression of expression of the alternatively spliced *FOXC2* gene, at either the transcriptional or translational level, is useful to generate cellular or animal models for diseases/conditions related to lipid metabolism.

20

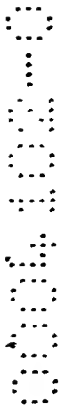
25

In yet another aspect, the invention provides a method for the identification of polypeptides which bind to nucleotide sequences involved in the biological pathway regulating lipid metabolism and/or adipocyte differentiation, comprising the steps of:

(a) transfecting a host cell line with a human *FOXC2* nucleotide sequence linked to a reporter gene, such as a gene encoding Green Fluorescent Protein (GFP) (for a review, see e.g. Galbraith et al. (1999) *Methods in Cell Biology* 58: 315-341);

30

(b) transfecting the said host cell line with a variety of human cDNA sequences, e.g. sequences included in a cDNA library;



- (c) identifying and isolating cells, e.g. by FACS cells sorting, having an altered level of expression of the said reporter gene, which is indicative that the polypeptide encoded by the added cDNA up- or downregulates at least one gene involved in the biological pathway regulating lipid metabolism and/or adipocyte differentiation;
- 5 (d) recovering cDNA from the cells isolated in step (c), by standard procedures, e.g. PCR or a CRE-LOX mediated procedure (see e.g. Sauer (1998) Methods 14: 381-392); and
- (e) identifying the polypeptide expressed by the cDNA recovered in step (d), e.g. by sequencing the cDNA and comparing the obtained sequence against sequence databases.
- 10 Throughout this description the terms "standard protocols" and "standard procedures", when used in the context of molecular biology techniques, are to be understood as protocols and procedures found in an ordinary laboratory manual such as: Current Protocols in Molecular Biology, editors F. Ausubel et al., John Wiley and Sons, Inc. 1994, or Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular Cloning: A laboratory manual,
- 15 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1989.

## EXAMPLES

- 20 Additional features of the invention will be apparent from the following Examples. Examples 1 to 4 are actual, while the remaining Examples are prophetic.

### EXAMPLE 1: Computational identification of *FOXC2* genomic sequences

- 25 The sequences present in the GenBank database (<http://www.ncbi.nlm.nih.gov>) were screened for sequence similarity to the human *FOXC2* cDNA sequence (GenBank accession number NM\_00521 (SEQ ID NO: 9)). The BLAST algorithm (Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402) was used for determining sequence identity. Software for performing BLAST analyses is publicly available through the National Center
- 30 for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). A working draft genomic sequence in 25 unordered pieces, from the *Homo sapiens* chromosome 16 clone RP11-46309 (GenBank accession number AC009108; Version 6; GI:7689930; released 4 May 2000), was selected for further studies.

Regions in sequence AC009108 matching portions of the *FOXC2* cDNA sequence NM\_005251 were combined using the PHRAP software, developed at the University of Washington (<http://www.genome.washington.edu/UWGC/analysistools/phrap.htm>). Two  
 5 contigs of 9780 bp (positions 116445 to 126224 in GenBank AC009108.6) and 3784 bp (positions 42927 to 46710 in GenBank AC0091108.6), respectively, were assembled to generate a human *FOXC2* genomic fragment of 13451 bp.

The ClustalW multiple sequence alignment program, version 1.8 (Thompson et al. (1994) Nucleic Acids Research 22: 4673-4680), was then used to identify the human *FOXC2*  
 10 extended genomic DNA sequence of 6458 bp (SEQ ID NO: 1) by comparison with the mouse cDNA sequence X74040 (SEQ ID NO: 6). First, a 6459 bp sequence, corresponding to positions 1500-7958 in the 13451 bp sequence, was selected. Positions 1-2285 in this 6459 bp sequence corresponded to 44426-46710 in AC009108.6, while positions 2151-  
 15 6459 corresponded to positions 126224-121916 (reverse complement taken) in AC009108.6. The overlap of positions 2151-2285 allowed for the contigs to be joined by the assembly program. The G residue in position 2655 was considered to be a sequencing error and was removed, which resulted in the 6458 bp sequence set forth as SEQ ID NO: 1. The open reading frame in SEQ ID NO: 1 encodes a polypeptide (SEQ ID NO: 2) identical  
 20 with the known human *FOXC2* polypeptide shown as SEQ ID NO:10.

#### EXAMPLE 2: Identification of potential regulatory sequences in the human and mouse *FOXC2* genomic sequences

25 In phylogenetic footprinting (for a review, see Duret & Bucher (1997) Current Opinion in Structural Biology 7(3): 399-406) sequences are aligned and a regional sequence identity is determined for each window of a fixed, arbitrary length. This allows the identification of potential regulatory regions in genomic sequences. Non-exon sequences that are conserved  
 30 over the course of evolution are likely to perform regulatory roles. Phylogenetic footprinting was performed as described in Wasserman & Fickett (1998) J. Mol. Biol. 278, 167-181, based on an alignment generated with the ClustalW multiple sequence alignment program, version 1.8 (Thompson et al. (1994) Nucleic Acids Research 22: 4673-4680),

with default parameters adjusted to a gap opening penalty of 20 and a gap extension penalty of 0.2. The human (SEQ ID NO: 1) and mouse (SEQ ID NO: 5) genomic sequences were aligned. Percentage identity was plotted for each contiguous 200 bp segment of the human gene to identify segments differentially conserved (in comparison to adjoining sequences) (Fig. 2).

In addition to segments of the published exon sequence, two differentially conserved regions or “footprints” were identified in the human gene. Both of these regions are local maxima and contain segments which exceed 70% nucleotide identity between the human and mouse genomic sequences. One region, shown as positions 1250 to 2235, in particular positions 1250 to 1749, in SEQ ID NO: 1, immediately adjacent to the published exon region, is likely to contain the transcription start site and proximal promoter regulatory sequences (Fig. 4). Another region, shown as positions 216 to 475 in SEQ ID NO: 1, approximately 1700 bp distal from the transcription start site, is likely to function as some form of regulatory region (either enhancer or repressor) (Fig. 3). (A schematic overview of the extended *FOXC2* gene is shown in Fig. 1.)

Further analysis of these regulatory regions identified short segments of higher conservation between the mouse and human genes, suggesting that these specific segments function as transcription factor binding sites. The *TRANSFAC* transcription factor database (<http://transfac.gbf.de>) (see Wingender et al. (2000) Nucleic Acids Research 28(1): 316-319) was screened for matches to known transcription factors. Consensus sites (identifiers R05066; R05067; R05068; and R05069) were found to match sequences conserved between the human *FOXC2* and mouse *FoxC2* genes. This suggests the presence of multiple forkhead-like binding sites in the distal regulatory enhancer, and potential auto-regulation of *FOXC2* by its protein product.

The same analysis was performed with reference to 200 bp contiguous segments of the mouse *FoxC2* genomic sequence (SEQ ID NO: 5). The following conserved regions were identified: 190 to 420; 1070 to 1645; and 5580 to 5875. They correlate to the regions indicated above for the human sequence and should be considered orthologous regions.

### EXAMPLE 3: Identification of an alternative human *FOXC2* cDNA sequence

BLASTN screening of the dbEST database from GenBank, using the human *FOXC2* cDNA (SEQ ID NO: 9) as a query sequence, revealed several ESTs overlapping containing portions of the available cDNA. A specialized tool, *est\_genome* (<http://www.sanger.ac.uk>), for the prediction of exon boundaries using ESTs was applied to compare the EST sequences to the genomic sequences (See Mott, R. (1997) Computer Applications in the Biosciences 13(4): 477-478). Two classes of ESTs were observed: sequences extending into the 3'-untranslated region and sequences revealing an alternative first exon spliced to a junction internal to the previously described first exon.

Specifically, it was found that the nucleotides in positions 33 to 182 in the EST with accession no. AW271272 (SEQ ID NO: 11) were identical to positions 66 to 215 in the extended *FOXC2* genomic sequence (SEQ ID NO: 1), and that positions 183 to 327 in SEQ ID NO: 11 were identical to positions 2516 to 2660 in SEQ ID NO: 1. Similarly, positions 5 to 55 in the EST with accession no. AW793237 (SEQ ID NO: 12) were identical to positions 165 to 215 in the extended *FOXC2* genomic sequence (SEQ ID NO: 1), and positions 56 to 157 in SEQ ID NO: 12 were identical to positions 2516 to 2607 in SEQ ID NO: 1. These results revealed an alternative splicing pattern in the human *FOXC2* gene. According to this splicing pattern, an alternative gene sequence (SEQ ID NO: 3) is derived by joining the regions shown as positions 1-215 and 2516-6458 in SEQ ID NO: 1. Alternative splicing patterns are known to regulate the synthesis of a variety of peptides and proteins. It may result in proteins with an entirely different function or in dysfunctional or inhibitory splice products (for a review, see McKeown (1992) Annu. Rev. Cell. Biol. 8: 133-155).

The amino acids corresponding to positions 1 to 94 in the published *FOXC2* transcription factor (SEQ ID NO: 10) are missing in protein encoded by the spliced variant generated from the alternative promoter (SEQ ID NO: 4). Consequently, the entire region N-terminal of the DNA binding domain and a portion of the DNA-binding domain (corresponding to positions 72-94 in SEQ ID NO: 2) are not present in the splice variant. It is postulated that this truncation leads to a protein which has a deficient "forkhead" DNA-binding region, and thus has a potential inhibitory function on the biological activities of the *FOXC2*

protein. This truncated FOXC2 protein may have a role in regulation of FOXC2, and an involvement in adipocyte differentiation and adipogenesis.

5    **EXAMPLE 4: Cloning and sequencing of the FOXC2 promoter**

The DNA region corresponding to nucleotide 176 to nucleotide 2233 (SEQ ID NO. 1 version 2) has been cloned using nested PCR on human genomic DNA. The PCR was performed according the Herculanse™ protocol (Stratagene catalog #600260;  
10    <http://www.stratagene.com/pcr/herculase.htm>) and with the inclusion of 8-10% DMSO.

In the initial reaction, the 5'-primer KRKX131 (CCATTGCCTTCTAGTCGCCTCC) was used together with the 3'-primer KRKX133 (CGTTGGGGTCGGACACGGAGTA) using 250 ng Clontech Genomic DNA # 6550-1 as template. The nested reaction was performed  
15    on 1/100 of the initial PCR reaction using the 5'-primer KRKX132 (GGTACCTACGCAGCCGATGAACAGCCA) and the 3'-primer KRKX134 (GCTAGCGCTGCTTCCGAGACGGCTCG). After the second PCR, the product was analyzed by electrophoresis in a 1.2% agarose gel, and a PCR product of the expected size was obtained and extracted for ligation into a TOPO PCR2.1 vector (Invitrogen, Carlsbad,  
20    CA) by standard cloning procedures and thereafter sequenced. The PCR reaction and cloning procedure was repeated in two parallel separate experiments, and sequence data from the two separate reactions were compared with the bioinformatically assembled sequence.

25    A DNA region containing the promoter (Fig. 4) corresponding to nt1179 to 2233 (SEQ ID NO: 1, version 2) was has been cloned using nested PCR in the same manner as described above. In the initial reaction, the 5'-primer KRKX136 (GGTACCCCCCGAGCCTGGAAACTCCCT) was used together with the 3'-primer KRKX134 (GCTAGCGCTGCTTCCGAGACGGCTCG) using 250 ng genomic DNA as a  
30    template. The PCR reaction and cloning procedure was repeated in four parallel separate experiments, and sequence data from the four separate reactions were compared with the bioinformatically assembled sequence.



**EXAMPLE 5: Tissue expression profiling of the alternative transcript**

Tissue expression profiling of the alternative transcript (SEQ ID NO: 3) is performed using standard Northern blotting procedures. RNA samples from an array of human tissues, including adipose tissue, are analyzed using an RNA or DNA probe specific for the alternative transcript. The expression profile in adipose tissue could be indicative a putative regulatory function for the alternative gene product (SEQ ID NO: 4) on adipogenesis and adipocyte differentiation.

In addition, reverse transcriptase PCR (RT-PCR) according to standard procedures is used to detect very low level expression of the alternative transcript in adipose tissue. RNA is prepared from human adipose tissue, and RT-PCR is performed using PCR primers specific for the alternative transcript.

**EXAMPLE 6: Mapping of the 5'-edge of the alternative exon by RACE-PCR**

RNA is prepared from human adipose tissue using standard protocols. RACE (Rapid Amplification of cDNA Ends) PCR is performed using the SMART<sup>TM</sup> RACE cDNA Amplification Kit (Clontech catalogue No. K1811-1; <http://www.clontech.com/product/catalog/PCR/smartrace.html>). With this procedure, the first strand synthesis produces cDNA with an extension containing a known sequence. Due to the mechanism of the extension procedure, the extension is typically added only to complete first strand cDNAs.

The 5'-RACE PCR is then performed using the 5'-primer provided with the kit, together with a 3'-primer corresponding to positions 210-237 in SEQ ID NO: 3 (GAACTGGTAGATGCCGTTCAAGGTTTCC) specific for the alternative transcript. The PCR product is cloned into a cloning vector and sequenced using standard protocols.

**EXAMPLE 7: Functional analysis**

The identified regulatory regions are analyzed to determine their impact on the transcription of the *FOXC2* gene or a reporter gene substituted for *FOXC2*. A PCR reaction is performed to isolate the promoter region adjacent to the published exon sequence, possibly including the sequences extending to the beginning of the ATG encoding the first methionine. This PCR product is cloned into a reporter plasmid adjacent to a reporter gene (e.g. luciferase). The upstream regulatory region, i.e. regions containing both upstream and promoter proximal sequences, or these sequences bearing artificially induced differences, are cloned in a similar manner. These constructs are transfected into a cell culture model system and the level/activity of the protein encoded by the reporter gene is determined. This would provide information on the function of the identified regions, and used to assess the impact of the different regions on transcriptional regulation. Similarly, the upstream regulatory region, a region containing both upstream and promoter proximal sequences, or these sequences bearing artificially induced differences can be cloned and used to assess the impact of these regions on the transcription of the reporter gene.

**EXAMPLE 8: Reporter gene assay to identify modulating compounds**

Reporter gene assays are well known as tools to signal transcriptional activity in cells. (For a review of chemiluminescent and bioluminescent reporter gene assays, see Bronstein et al. (1994) Analytical Biochemistry 219, 169-181.) For instance, the photoprotein luciferase provides a useful tool for assaying for modulators of promoter activity. Cells are transiently transfected with a reporter construct which includes a gene for the luciferase protein downstream from the *FOXC2* promoter and enhancer region, or fragments thereof regulating the *FOXC2* activity. Luciferase activity may be quantitatively measured using e.g. luciferase assay reagents that are commercially available from Promega (Madison, WI). Differences in luminescence in the presence versus the absence of a candidate modulator compound are indicative of modulatory activity.

TABLE I

Summary of *FOXC2* sequences

SEQ ID NO:	GenBank accession no.	Description
1		Human <i>FOXC2</i> extended genomic DNA sequence
2		Human <i>FOXC2</i> polypeptide sequence (Identical with SEQ ID NO: 10)
3		Human <i>FOXC2</i> DNA sequence Alternative splicing
4		Human polypeptide sequence Alternative open reading frame
5	Y08222	Mouse MHF-1 ( <i>FoxC2</i> ) genomic DNA sequence (CDS 2070 – 3554)
6	X74040	Mouse MHF-1 ( <i>FoxC2</i> ) cDNA sequence
7		Mouse MHF-1 ( <i>FoxC2</i> ) polypeptide sequence
8	Y08223	Human <i>FKHL14</i> ( <i>FOXC2</i> ) genomic DNA sequence (CDS 1197 – 2702)
9	NM_005251	Human <i>FKHL14</i> ( <i>FOXC2</i> ) cDNA sequence
10		Human <i>FKHL14</i> ( <i>FOXC2</i> ) polypeptide sequence
11	AW 271272	Human EST
12	AW 793237	Human EST

0  
 1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12

TABLE II

Summary of features in human *FOXC2* sequences shown as SEQ ID NOs: 1 and 3

Feature	Positions
SEQ ID NO: 1	
First exon according to the alternative transcript	1 – 215
– Untranslated region	1 – 186
– Region coding for 5'-part of alternative protein	187 – 215
Alternative first exon splice site	215 – 216
Predicted enhancer region	216 – 475
– E-box-like region	223 – 231
– Forkhead-like region	359 – 375
– Forkhead-like region	378 – 402
– Ets-like region	403 – 423
Predicted promoter region	1250 – 1749
– Forkhead-like region	1692 – 1703
First exon according to the published form of the transcript	1746 – 4629
– Untranslated region	1746 – 2234
– Polypeptide coding region	2235 – 3740
– Region coding for DNA-binding domain	2448 – 2735
Second exon according to the alternative transcript	2516 – 4629
– Portion of polypeptide used in alternative transcript	2516 – 3740
– Untranslated region	3741 – 4629
SEQ ID NO: 3	
Polypeptide coding region (5' of splice site)	187 – 215
Polypeptide coding region (3' of splice site)	216 – 1437
– Region coding for truncated portion of protein	216 – 435

## CLAIMS

1. A human *FOXC2* promoter region comprising a sequence selected from:
  - (a) the nucleotide sequence set forth as positions 1692 to 1703 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* promoter activity;
  - (b) the complementary strand of (a); and
  - (c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).
2. The human *FOXC2* promoter region according to claim 1, comprising a sequence selected from:
  - (a) the nucleotide sequence set forth as positions 1250 to 1749 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* promoter activity;
  - (b) the complementary strand of (a); and
  - (c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).
3. The human *FOXC2* promoter region according to claim 2, comprising a sequence selected from:
  - (a) the nucleotide sequence set forth as positions 1250 to 2235 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* promoter activity;
  - (b) the complementary strand of (a); and
  - (c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).
4. A recombinant construct comprising the human *FOXC2* promoter region according to any one of claims 1 to 3.
5. The recombinant construct according to claim 4 wherein the human *FOXC2* promoter region is operably linked to a gene encoding a detectable product.
6. The recombinant construct according to claim 5 wherein said gene encoding a detectable product is a human *FOXC2* gene.

7. The recombinant construct according to claim 4 further comprising a reporter gene.
8. A vector comprising the recombinant construct according to any one of claims 4 to 7.
- 5 9. A host cell stably transformed with the recombinant construct according to any one of claims 4 to 7.
- 10 10. A method for identification of an agent regulating *FOXC2* promoter activity, said method comprising the steps
- (i) contacting a candidate agent with a human *FOXC2* promoter region as defined in any one of claims 1 to 3; and
- (ii) determining whether said candidate agent modulates expression of the *FOXC2* gene, such modulation being indicative for an agent capable of regulating *FOXC2* promoter activity.
- 15 11. A method for identification of an agent regulating *FOXC2* promoter activity, said method comprising assaying reporter gene expression in a cell according to claim 9 in the presence and absence of a candidate agent, wherein an effect on the test level
- 20 of expression as compared to control level of expression is indicative of an agent capable of regulating *FOXC2* promoter activity.
12. A human *FOXC2* enhancer region comprising a sequence selected from:
- (a) the nucleotide sequence set forth as positions 223 to 231 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;
- 25 (b) the complementary strand of (a); and
- (c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).
- 30 13. A human *FOXC2* enhancer region comprising a sequence selected from:
- (a) the nucleotide sequence set forth as positions 359 to 375 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;
- (b) the complementary strand of (a); and

(c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).

14. A human *FOXC2* enhancer region comprising a sequence selected from:

(a) the nucleotide sequence set forth as positions 378 to 402 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;

(b) the complementary strand of (a); and

(c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).

15. A human *FOXC2* enhancer region comprising a sequence selected from:

(a) the nucleotide sequence set forth as positions 403 to 423 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;

(b) the complementary strand of (a); and

(c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).

16. The human *FOXC2* enhancer region according to any one of claims 12 to 15 comprising a sequence selected from:

(a) the nucleotide sequence set forth as positions 216 to 475 in SEQ ID NO: 1, or a fragment thereof exhibiting *FOXC2* enhancer activity;

(b) the complementary strand of (a); and

(c) nucleotide sequences capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence as defined in (a) or (b).

17. A recombinant construct comprising a human *FOXC2* enhancer region according to any one of claims 12 to 15.

18. A vector comprising the recombinant construct according to claim 17.

19. A host cell stably transformed with the recombinant construct according to claim 18.

20. A method for identification of an agent regulating *FOXC2* enhancer activity, said method comprising the steps

(i) contacting a candidate agent with the human *FOXC2* enhancer region as defined in any one of claims 12 to 16; and

5 (ii) determining whether said candidate agent modulates expression of the *FOXC2* gene, such modulation being indicative for an agent capable of regulating *FOXC2* enhancer activity.

21. A method for identification of an agent capable of regulating *FOXC2* enhancer activity, said method comprising assaying reporter gene expression in a cell as defined in claim 19 in the presence and absence of a candidate agent, wherein an effect on the test level of expression as compared to control level of expression is indicative of an agent capable of regulating *FOXC2* enhancer activity.

10

22. A method for identification of an agent capable of regulating a mammalian *FOXC2* promoter activity, said method comprising the steps

(i) contacting a candidate agent with a murine *FoxC2* promoter nucleotide sequence shown as positions 1250 to 2235 in SEQ ID NO: 5; and

(ii) determining whether said candidate agent modulates expression of a mammalian *FOXC2* gene, such modulation being indicative for an agent capable of regulating mammalian *FOXC2* promoter activity.

15 20

23. A method for identification of an agent capable of regulating a mammalian *FOXC2* enhancer activity, said method comprising the steps

(i) contacting a candidate agent with a murine *FoxC2* enhancer nucleotide sequence shown as positions 216 to 475 in SEQ ID NO: 5; and

(ii) determining whether said candidate agent modulates expression of a mammalian *FOXC2* gene, such modulation being indicative for an agent capable of regulating mammalian *FOXC2* enhancer activity.

25 30

24. A method for identification of an agent capable of regulating a mammalian *FOXC2* enhancer activity, said method comprising the steps



(i) contacting a candidate agent with a murine *FoxC2* enhancer nucleotide sequence shown as positions 216 to 2235 in SEQ ID NO: 5; and

(ii) determining whether said candidate agent modulates expression of a mammalian *FOXC2* gene, such modulation being indicative for an agent capable of regulating mammalian *FOXC2* enhancer activity.

25. An isolated nucleic acid molecule selected from:

(a) nucleic acid molecules comprising a nucleotide sequence as shown in SEQ ID NO: 3;

(b) nucleic acid molecules comprising a nucleotide sequence capable of hybridizing, under stringent hybridization conditions, to a nucleotide sequence complementary the polypeptide coding region of a nucleic acid molecule as defined in (a) and which codes for a variant form of the FOXC2 transcription factor; and

(c) nucleic acid molecules comprising a nucleic acid sequence which is degenerate as a result of the genetic code to a nucleotide sequence as defined in (a) or (b) and which codes for a variant form of the FOXC2 transcription factor.

26. An isolated polypeptide encoded by the nucleic acid according to claim 25.

27. The isolated polypeptide according to claim 26 having an amino acid sequence shown as SEQ ID NO: 4 in the Sequence Listing

28. A vector harboring the nucleic acid molecule according to claim 25.

29. A replicable expression vector, which carries and is capable of mediating the expression of a nucleotide sequence according to claim 25.

30. A cultured host cell harboring a vector according to claim 28 or 29.

31. A process for production of a variant form of the FOXC2 transcription factor polypeptide, comprising culturing a host cell according to claim 30 under conditions whereby said polypeptide is produced, and recovering said polypeptide.

32. A method for identifying an agent capable of regulating expression of the nucleic acid molecule according to claim 25, said method comprising the steps
- (i) contacting a candidate agent with the said nucleic acid molecule; and
- (ii) determining whether said candidate agent modulates expression of the said nucleic acid molecule.
33. An antisense oligonucleotide having a sequence capable of specifically hybridizing to RNA transcribed by the nucleic acid molecule according to claim 25, so as to prevent translation of the said RNA.
34. A method for the identification of polypeptides which bind to nucleotide sequences involved in the biological pathway regulating lipid metabolism and/or adipocyte differentiation, comprising
- (a) transfecting a host cell line with a human FOXC2 nucleotide sequence linked to a reporter gene;
- (b) transfecting the said host cell line with a variety of human cDNA sequences;
- (c) identifying and isolating cells having an altered level of expression of the said reporter gene;
- (d) recovering cDNA from the cells isolated in step (c); and
- (e) identifying the polypeptide expressed by the cDNA recovered in step (d).

## ABSTRACT

The present invention relates an isolated promoter region of the mammalian transcription factor *FOXC2*. The invention also relates to screening methods for agents modulating the expression of *FOXC2* and thereby being potentially useful for the treatment of medical conditions related to obesity. The invention further relates to a previously unknown variant  
5 of the human *FOXC2* gene, derived via the use of an alternative promoter, which produces an additional exon that generates a distinct open reading frame via splicing. The alternative gene encodes a variant of the FOXC2 transcription factor, which is lacking a part of the DNA-binding domain and consequently has a potential regulatory function.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10

- 1 -

## SEQUENCE LISTING

&lt;110&gt; Pharmacia &amp; Upjohn AB

&lt;120&gt; Promoter Sequences

&lt;130&gt; 00298

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 12

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 6458

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (2235)..(3740)

&lt;400&gt; 1

```

cctttggcctt tgaattgatac aggagacaaa gataatgcat ctacattttc gtcttctgtt 60
cttttattgg aaataagtgg cagcctccat tgccttctag tcgcctcccc gaagcgaaga 120
ggccgaagcg aagaggcctg gtgggtgtgc tcaacatcct tttgctgaga atcgaatacg 180
cagccgatga acagccagga aggggtgcaag gaaacctgaa atacaaatgt tctccctgaa 240
gccctcttcc ctgcccaccc agaccagcaa ctcccaaat tctgcccgtg tttagccttg 300
ttaaaggggt gtctcactcc ttcagggaag gtgggaaaag gggatctgat tattgagggtg 360
tggaaggaat aaataatcag tccacaaata acaaaactgt ccgggattcc tagagggaag 420
gagaaatcct tgaaggagat ccaagtgcct ccaggctctgc ctgccgaata atatcatccc 480
gaagggatct tgaaccgttt gcaatcaacc gtcacccag tcttccacg gagcgcgctc 540
cctaactcac cctaccacc caacaaaaca aaaaaaggc tgaaatatag aaaagcaact 600
tggaggctcc cagggggacg ttgccaggag caggaggcag ggacagcgcc ctagggtcgg 660
tgttagcggc cggcgccggc ctgggccacg ggaaacgtcc acgcttggtg cccgcggtgc 720
gcggcgctca ttgcgcgcgc cttcgagcca agccccgcg gaaaacaggc tcgggtttct 780
cctcgcaggg cccaggaact cggctctgcc tggcccggtt gggctcgtgc attgtcccgg 840
tcttctggga gtgcgggggc agcttggttag agggaaattc tacctgggaa aaggagagcg 900
agtttcgaag ctgaagttgg taggctgcga gtgtccacgc gggagacgaa agggggaaat 960
agcagagtca cttcaccctt ttcccaaac cccacaaaac tgctcgcagc gacgcggatg 1020

```

- 2 -

atctaccgaa ttccccgcga attcggagga ttaagttgtc agtcagcacg ttgtacctt 1080  
 cccctctatg cactccgctg cctggctcct cggcggggag cgagggaaac tcagtttgta 1140  
 gggtttacct ctaaaacotc gataggttat ccttgacgac cccgagcctg gaaactccct 1200  
 gttgatgatt aattatttga ttaaataagt ataacatcca ggagaggccc tgccattcca 1260  
 atccagcgcg tttgcttttg aatccattac acctgggccc ccataattag gaaatctaata 1320  
 tattcgcttc atcactcatt aataagaaaa atgtcccagg atcattgcta cttacaagggt 1380  
 ctttgggaga gatattttac tctattaatc cattctattt tatatttcaa attgattttt 1440  
 tttacagag gaaagtggct atctttttgt tttgggcatg tgggcccatt caccaaaatg 1500  
 tgatcataaa ataaatttta ataagatata actttttaaa aagttttcaa gtgaagacgg 1560  
 agtcgcccgc gaggccgggg cggcggggtc ttagagccga cggattcctg cgctcctcgc 1620  
 cccgattggc gccggactcc tctcagctgc cgggtgattg gctcaaagtt ccgggagggg 1680  
 gcgtggcccc aggaaagtaa aaactcgctt tcagcaagaa gacttttgaa acttttccca 1740  
 atccctaaaa gggacttggc ctctttttct gggctcagcg gggcagccgc tcggaccccc 1800  
 gcgcgctgac cctcggggct gccgattcgc tgggggcttg gagagcctcc tgcgccctc 1860  
 ctgcgcggg ccgaggggtc accttgggtc ccaggccgcg gcgtctccgc tgggtccgcg 1920  
 gccgcccgc tgcgcgcgt gccgcgcgc ggtcctggag ccagcgagga gcggggccgg 1980  
 cgctgcgctt gcccggggcg cgccctccag gatgccgatc cgcccggtcc gctgaaagcg 2040  
 cgccccctg ctcgcccga gcgacgacga ccgcgcaccc tcgccccga ggctgccagg 2100  
 agaccggggc cgccctccc gctccctcc tctccctc tggctctctc gcgctctctc 2160  
 gctctcaggg cccctcgc tccccggcc gcagtccgtg cgcgagggcg ccggcgagcc 2220  
 gtctcggaag cagc atg cag gcg cgc tac tcc gtg tcc gac ccc aac gcc 2270  
                   Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala  
                   1                  5                  10  
 ctg gga gtg gtg ccc tac ctg agc gag cag aat tac tac cgg gct gcg 2318  
 Leu Gly Val Val Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala  
                   15                  20                  25  
 ggc agc tac ggc ggc atg gcc agc ccc atg ggc gtc tat tcc ggc cac 2366  
 Gly Ser Tyr Gly Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His  
                   30                  35                  40  
 ccg gag cag tac agc gcg ggg atg ggc cgc tcc tac gcg ccc tac cac 2414  
 Pro Glu Gln Tyr Ser Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His  
                   45                  50                  55                  60  
 cac cac cag ccc gcg gcg cct aag gac ctg gtg aag ccg ccc tac agc 2462  
 His His Gln Pro Ala Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser  
                   65                  70                  75

- 3 -

tac atc gcg ctc atc acc atg gcc atc cag aac gcg ccc gag aag aag	2510
Tyr Ile Ala Leu Ile Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys	
80 85 90	
atc acc ttg aac ggc atc tac cag ttc atc atg gac cgc ttc ccc ttc	2558
Ile Thr Leu Asn Gly Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe	
95 100 105	
tac cgg gag aac aag cag ggc tgg cag aac agc atc cgc cac aac ctc	2606
Tyr Arg Glu Asn Lys Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu	
110 115 120	
tgc ctc aac gag tgc ttc gtc aag gtg ccc cgc gac gac aag aag ccc	2654
Ser Leu Asn Glu Cys Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro	
125 130 135 140	
ggc aag ggc agt tac tgg acc ctg gac ccg gac tcc tac aac atg ttc	2702
Gly Lys Gly Ser Tyr Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe	
145 150 155	
gag aac ggc agc ttc ctg cgg cgc cgg cgc ttc aaa aag aag gac	2750
Glu Asn Gly Ser Phe Leu Arg Arg Arg Arg Phe Lys Lys Lys Asp	
160 165 170	
gtg tcc aag gag aag gag gag cgg gcc cac ctc aag gag ccg ccc ccg	2798
Val Ser Lys Glu Lys Glu Glu Arg Ala His Leu Lys Glu Pro Pro Pro	
175 180 185	
gcg gcg tcc aag ggc gcc ccg gcc acc ccc cac cta gcg gac gcc ccc	2846
Ala Ala Ser Lys Gly Ala Pro Ala Thr Pro His Leu Ala Asp Ala Pro	
190 195 200	
aag gag gcc gag aag aag gtg gtg atc aag agc gag gcg gcg tcc ccg	2894
Lys Glu Ala Glu Lys Lys Val Val Ile Lys Ser Glu Ala Ala Ser Pro	
205 210 215 220	
gcg ctg ccg gtc atc acc aag gtg gag acg ctg agc ccc gag agc gcg	2942
Ala Leu Pro Val Ile Thr Lys Val Glu Thr Leu Ser Pro Glu Ser Ala	
225 230 235	
ctg cag ggc agc ccg cgc agc gcg gcc tcc acg ccc gcc ggc tcc ccc	2990
Leu Gln Gly Ser Pro Arg Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro	
240 245 250	
gac ggt tcg ctg ccg gag cac cac gcc gcg gcg ccc aac ggg ctg cct	3038
Asp Gly Ser Leu Pro Glu His His Ala Ala Ala Pro Asn Gly Leu Pro	
255 260 265	
ggc ttc agc gtg gag aac atc atg acc ctg cga acg tcg ccg ccg ggc	3086
Gly Phe Ser Val Glu Asn Ile Met Thr Leu Arg Thr Ser Pro Pro Gly	
270 275 280	
gga gag ctg agc ccg ggg gcc gga cgc gcg ggc ctg gtg gtg ccg ccg	3134
Gly Glu Leu Ser Pro Gly Ala Gly Arg Ala Gly Leu Val Val Pro Pro	
285 290 295 300	

- 4 -

ctg gcg ctg cca tac gcc gcc gcg ccg ccc gcc gcc tac ggc cag ccg 3182  
 Leu Ala Leu Pro Tyr Ala Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro  
 305 310 315

tgc gct cag ggc ctg gag gcc ggg gcc gcc ggg ggc tac cag tgc agc 3230  
 Cys Ala Gln Gly Leu Glu Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser  
 320 325 330

atg cga gcg atg agc ctg tac acc ggg gcc gag cgg ccg gcg cac atg 3278  
 Met Arg Ala Met Ser Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Met  
 335 340 345

tgc gtc ccg ccc gcc ctg gac gag gcc ctc tcg gac cac ccg agc ggc 3326  
 Cys Val Pro Pro Ala Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly  
 350 355 360

ccc acg tcg ccc ctg agc gct ctc aac ctc gcc gcc ggc cag gag ggc 3374  
 Pro Thr Ser Pro Leu Ser Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly  
 365 370 375 380

gcg ctc gcc gcc acg ggc cac cac cac cag cac cac ggc cac cac cac 3422  
 Ala Leu Ala Ala Thr Gly His His His Gln His His Gly His His His  
 385 390 395

ccg cag gcg ccg ccg ccc ccg ccg gct ccc cag ccc cag ccg acg ccg 3470  
 Pro Gln Ala Pro Pro Pro Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro  
 400 405 410

cag ccc ggg gcc gcc gcg gcg cag gcg gcc tcc tgg tat ctc aac cac 3518  
 Gln Pro Gly Ala Ala Ala Ala Gln Ala Ala Ser Trp Tyr Leu Asn His  
 415 420 425

agc ggg gac ctg aac cac ctc ccc ggc cac acg ttc gcg gcc cag cag 3566  
 Ser Gly Asp Leu Asn His Leu Pro Gly His Thr Phe Ala Ala Gln Gln  
 430 435 440

caa act ttc ccc aac gtg cgg gag atg ttc aac tcc cac cgg ctg ggg 3614  
 Gln Thr Phe Pro Asn Val Arg Glu Met Phe Asn Ser His Arg Leu Gly  
 445 450 455 460

att gag aac tcg acc ctc ggg gag tcc cag gtg agt ggc aat gcc agc 3662  
 Ile Glu Asn Ser Thr Leu Gly Glu Ser Gln Val Ser Gly Asn Ala Ser  
 465 470 475

tgc cag ctg ccc tac aga tcc acg ccg cct ctc tat cgc cac gca gcc 3710  
 Cys Gln Leu Pro Tyr Arg Ser Thr Pro Pro Leu Tyr Arg His Ala Ala  
 480 485 490

ccc tac tcc tac gac tgc acg aaa tac tga cgtgtcccg gacctccct 3760  
 Pro Tyr Ser Tyr Asp Cys Thr Lys Tyr  
 495 500

ccccggcccg ctccggtctt gcttcccagc cccgacccaa ccagacaatt aaggggctgc 3820

agagacgcaa aaaagaaaca aaacatgtcc accaaccctt tctcagaccc gggagcagag 3880

agcgggcacg ctagccccca gccgtctgtg aagagcgcag gtaactttta ttcgccgccc 3940

cgtttctggg atcccaggaa acccctccaa agggacgcag cccaacaaaa tgagtattgg 4000

- 5 -

tcttaaaatc cccctcccct accaggacgg ctgtgctgtg ctcgacctga gctttcaaaa 4060  
gttaagttat ggacccaaat cccatagcga gccctagtgt actttctgta ggggtcccca 4120  
taggtgtatg ggggtctcta tagataatat atgtgctgtg tgtaatttta aatttctcca 4180  
accgtgctgt acaaagtgtg ggatttgtta tcaggctatt ttgttgttgt tgttgttgtt 4240  
cagagccatt aatataatat ttaaagttga gttcactgga taagtttttc atcttgccca 4300  
accatttcta actgccaaat tgaattcaag aaaccgatgt gggttttgtt tctgtacaa 4360  
ttatgagata taattctttt tcccattgta ggtcttttac aaaacaagaa aataatttat 4420  
ttttttgttg gtggataaag aagtcaagta tctgatactt tttatttaca aagtgtgatg 4480  
gttttgtata gtaggttcca ccctgagtat tcctaaaaga aaaaaaaaaa aaaagcttaa 4540  
aaactctaac ttcattctgtg tttgtcttac gtggctctta tcgttgtact taccttaaaa 4600  
taaaccatg ttgttttttc tgcccaaagt ttggacagtg tgtttgtgtt gttgcatttt 4660  
ttacaaacga ggtgtgtttg caaacccacc tgctttgatt atttttgtta cacagggtggg 4720  
tatatgtgta gacacataaa aacgaccaga gaataggagc acacacctgc tgtcttgttt 4780  
agtgcagaaa aaaggctttt gattaatttt aaaatccac tctaggattt tttcttttcg 4840  
agaaaccgcc cagttggagg gggctgcctg aaggaccgga ccatgagttt gccgtgatgc 4900  
attttcttaa atgcacaaaa acatgctaata tgtcaaaaca aacagtgccca ctccatctca 4960  
gtgtccagcc gtccccagtt taggaggtga aggaagggaa gaataaacat ttcccgtttg 5020  
ctaactgcaa ccagggtga gtctgtctt ccccgatth tataaaattt gagcctcttt 5080  
gcctgtctta atagttttcc agagaatttg aactgggcca atgaaggtct gaaggggacg 5140  
gattttctag cgtttgatat ccatccccct tagcggccag atcagagggg aatttcagac 5200  
tttattactt ctcaatgtca tgtctaaatc tacaccctca tcgcagtga aaattttaaa 5260  
acctcattac ccttcaaaaa taatttatga tatttttaga gttctaaatt caagtttttc 5320  
aatatgttaa ataatagaga ttattttttg ttttcaatgt taatatctcg tcttttacat 5380  
ttttaatagt aacatagttt ttgtgaaatg tagctgacga aatggcttta ttatctattt 5440  
caatggctga agtccaccac tcccctgctg gcctctatgt gtgaatttgg ggaccaaagc 5500  
ttcatcaatt ccacccccag caggtgagct gtaccttgc aatgctgaag ttctttgtga 5560  
gcttaacggt tcaagaccag atgattttgc taaaggtgat ttgcttgat gcagtggcgc 5620  
tgaacgtaac cgggtgttt ttgtcgtgtt gttttcaaca tggcacttta tctccacgct 5680  
atgttgaaat agaattaggg gaagcttaaa gcataataat tgtccccaca tgtgcaacac 5740



- 6 -

agactctttc aatctgtggc cccagaggtg gcacacagtt aagacttggc ggctgtctca 5800  
 ttctttttca taatgtgagg gttcccgggt gtccgggtgc tagactttca gcaggcccca 5860  
 ggccagacgg gctttggttg agtgaacagg aggaggaagt taaggaggta ggggtgggga 5920  
 gagacctct ccaagctgca gaagaaggtg gcccaagctc cttgcctgag tctgccgtga 5980  
 tggtttcatt ttacttctgc tcgcttcacg ctatttgccc caggagaaga ggagagtatt 6040  
 ccagacggta agcgagctgg ctttttcct tccctagacg ttttaaagaa atctttctga 6100  
 aagcttgccc tcatcgtaag ctttgaaacc gttggtgtcc tgtagtgagg gagggctgag 6160  
 agacacggc agaaataaag gagagcgacg gtgtggctga gagccccag gtctgtgtt 6220  
 gaaactaagc tgggcttttg cacctttagg aagccttttt aaagaagtcc tgctgtgtgg 6280  
 gggccggaag cccaagttag tgggccttgc ggaggttatc gggaggggtc ttaccactc 6340  
 cttggggaac gtgggcaacg gggggattgt atctgaagct ttattcaggt cttcggcggc 6400  
 agcagagtgg agaaccaggc ccttagtggt tagcggcctg gggattttgg gactcacc 6458

&lt;210&gt; 2

&lt;211&gt; 501

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

Met	Gln	Ala	Arg	Tyr	Ser	Val	Ser	Asp	Pro	Asn	Ala	Leu	Gly	Val	Val
1				5					10					15	
Pro	Tyr	Leu	Ser	Glu	Gln	Asn	Tyr	Tyr	Arg	Ala	Ala	Gly	Ser	Tyr	Gly
			20					25					30		
Gly	Met	Ala	Ser	Pro	Met	Gly	Val	Tyr	Ser	Gly	His	Pro	Glu	Gln	Tyr
		35				40					45				
Ser	Ala	Gly	Met	Gly	Arg	Ser	Tyr	Ala	Pro	Tyr	His	His	His	Gln	Pro
	50				55						60				
Ala	Ala	Pro	Lys	Asp	Leu	Val	Lys	Pro	Pro	Tyr	Ser	Tyr	Ile	Ala	Leu
	65				70					75					80
Ile	Thr	Met	Ala	Ile	Gln	Asn	Ala	Pro	Glu	Lys	Lys	Ile	Thr	Leu	Asn
				85					90					95	
Gly	Ile	Tyr	Gln	Phe	Ile	Met	Asp	Arg	Phe	Pro	Phe	Tyr	Arg	Glu	Asn
		100					105						110		
Lys	Gln	Gly	Trp	Gln	Asn	Ser	Ile	Arg	His	Asn	Leu	Ser	Leu	Asn	Glu
		115				120						125			
Cys	Phe	Val	Lys	Val	Pro	Arg	Asp	Asp	Lys	Lys	Pro	Gly	Lys	Gly	Ser
	130					135					140				
Tyr	Trp	Thr	Leu	Asp	Pro	Asp	Ser	Tyr	Asn	Met	Phe	Glu	Asn	Gly	Ser
	145			150					155						160
Phe	Leu	Arg	Arg	Arg	Arg	Arg	Phe	Lys	Lys	Lys	Asp	Val	Ser	Lys	Glu
			165					170						175	
Lys	Glu	Glu	Arg	Ala	His	Leu	Lys	Glu	Pro	Pro	Pro	Ala	Ala	Ser	Lys
		180					185						190		
Gly	Ala	Pro	Ala	Thr	Pro	His	Leu	Ala	Asp	Ala	Pro	Lys	Glu	Ala	Glu
		195					200					205			
Lys	Lys	Val	Val	Ile	Lys	Ser	Glu	Ala	Ala	Ser	Pro	Ala	Leu	Pro	Val
	210					215						220			

- 7 -

```

Ile Thr Lys Val Glu Thr Leu Ser Pro Glu Ser Ala Leu Gln Gly Ser
225      230      235      240
Pro Arg Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu
      245      250      255
Pro Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val
      260      265      270
Glu Asn Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser
      275      280      285
Pro Gly Ala Gly Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro
      290      295      300
Tyr Ala Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly
305      310      315      320
Leu Glu Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met
      325      330      335
Ser Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro
      340      345      350
Ala Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro
      355      360      365
Leu Ser Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala
      370      375      380
Thr Gly His His His Gln His His Gly His His His Pro Gln Ala Pro
385      390      395      400
Pro Pro Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala
      405      410      415
Ala Ala Ala Gln Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu
      420      425      430
Asn His Leu Pro Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro
      435      440      445
Asn Val Arg Glu Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser
      450      455      460
Thr Leu Gly Glu Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro
465      470      475      480
Tyr Arg Ser Thr Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr
      485      490      495
Asp Cys Thr Lys Tyr
      500

```

```

<210> 3
<211> 4158
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> (187)..(1437)

```

```

<400> 3
cctttggcctt tgaattgcatc aggagacaaa gataatgcat ctacattttc gtcttctgtt 60
cttttatttgg aaataagtgg cacgccccat tgccttctag tcgcctcccc gaagcgaaga 120
ggccgaagcg aagaggcctg gtgggttgtc tcaacatcct tttgctgaga atcgaatacg 180
cagccg atg aac agc cag gaa ggg tgc aag gaa acc ttg aac ggc atc      228
Met Asn Ser Gln Glu Gly Cys Lys Glu Thr Leu Asn Gly Ile
      1              5              10

```

- 8 -

tac cag ttc atc atg gac cgc ttc ccc ttc tac cgg gag aac aag cag	276
Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn Lys Gln	
15 20 25 30	
ggc tgg cag aac agc atc cgc cac aac ctc tcg ctc aac gag tgc ttc	324
Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu Cys Phe	
35 40 45	
gtc aag gtg ccc cgc gac gac aag aag ccc ggc aag ggc agt tac tgg	372
Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser Tyr Trp	
50 55 60	
acc ctg gac ccg gac tcc tac aac atg ttc gag aac ggc agc ttc ctg	420
Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser Phe Leu	
65 70 75	
cgg cgc cgg cgg cgc ttc aaa aag aag gac gtg tcc aag gag aag gag	468
Arg Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Ser Lys Glu Lys Glu	
80 85 90	
gag cgg gcc cac ctc aag gag ccg ccc ccg gcg gcg tcc aag ggc gcc	516
Glu Arg Ala His Leu Lys Glu Pro Pro Pro Ala Ala Ser Lys Gly Ala	
95 100 105 110	
ccg gcc acc ccc cac cta gcg gac gcc ccc aag gag gcc gag aag aag	564
Pro Ala Thr Pro His Leu Ala Asp Ala Pro Lys Glu Ala Glu Lys Lys	
115 120 125	
gtg gtg atc aag agc gag gcg gcg tcc ccg gcg ctg ccg gtc atc acc	612
Val Val Ile Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val Ile Thr	
130 135 140	
aag gtg gag acg ctg agc ccc gag agc gcg ctg cag ggc agc ccg cgc	660
Lys Val Glu Thr Leu Ser Pro Glu Ser Ala Leu Gln Gly Ser Pro Arg	
145 150 155	
agc gcg gcc tcc acg ccc gcc ggc tcc ccc gac ggt tcg ctg ccg gag	708
Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu Pro Glu	
160 165 170	
cac cac gcc gcg gcg ccc aac ggg ctg cct ggc ttc agc gtg gag aac	756
His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val Glu Asn	
175 180 185 190	
atc atg acc ctg cga acg tcg ccg ccg ggc gga gag ctg agc ccg ggg	804
Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser Pro Gly	
195 200 205	
gcc gga cgc gcg ggc ctg gtg gtg ccg ccg ctg gcg ctg cca tac gcc	852
Ala Gly Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro Tyr Ala	
210 215 220	
gcc gcg ccg ccc gcc gcc tac ggc cag ccg tgc gct cag ggc ctg gag	900
Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly Leu Glu	
225 230 235	
gcc ggg gcc gcc ggg ggc tac cag tgc agc atg cga gcg atg agc ctg	948
Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met Ser Leu	
240 245 250	

- 9 -

tac acc ggg gcc gag cgg ccg gcg cac atg tgc gtc ccg ccc gcc ctg 996  
 Tyr Thr Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro Ala Leu  
 255 260 265 270

gac gag gcc ctc tcg gac cac ccg agc ggc ccc acg tcg ccc ctg agc 1044  
 Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro Leu Ser  
 275 280 285

gct ctc aac ctc gcc gcc ggc cag gag ggc gcg ctc gcc gcc acg ggc 1092  
 Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala Thr Gly  
 290 295 300

cac cac cac cag cac cac ggc cac cac cac ccg cag gcg ccg ccg ccc 1140  
 His His His Gln His His Gly His His His Pro Gln Ala Pro Pro Pro  
 305 310 315

ccg ccg gct ccc cag ccc cag ccg acg ccg cag ccc ggg gcc gcc gcg 1188  
 Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala Ala Ala  
 320 325 330

gcg cag gcg gcc tcc tgg tat ctc aac cac agc ggg gac ctg aac cac 1236  
 Ala Gln Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu Asn His  
 335 340 345 350

ctc ccc ggc cac acg ttc gcg gcc cag cag caa act ttc ccc aac gtg 1284  
 Leu Pro Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro Asn Val  
 355 360 365

ccg gag atg ttc aac tcc cac ccg ctg ggg att gag aac tcg acc ctc 1332  
 Arg Glu Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser Thr Leu  
 370 375 380

ggg gag tcc cag gtg agt ggc aat gcc agc tgc cag ctg ccc tac aga 1380  
 Gly Glu Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro Tyr Arg  
 385 390 395

tcc acg ccg cct ctc tat cgc cac gca gcc ccc tac tcc tac gac tgc 1428  
 Ser Thr Pro Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr Asp Cys  
 400 405 410

acg aaa tac tgacgtgtcc cgggacctcc cctccccggc ccgctccggc 1477  
 Thr Lys Tyr  
 415

ttgcgttccc agccccgacc caaccagaca attaaggggc tgcagagacg caaaaaagaa 1537

acaaaacatg tccaccaacc tttttctcaga cccgggagca gagagcgggc acgctagccc 1597

ccagccgtct gtgaagagcg caggtaactt taattcgccg ccccgtttct gggatcccag 1657

gaaacccctc caaagggacg cagcccaaca aaatgagtat tggctctaaa atccccctcc 1717

cctaccagga cggctgtgct gtgctcgacc tgagctttca aaagttaagt tatggacca 1777

aatcccatag cgagccccta gtgactttct gtaggggtcc ccataggtgt atgggggtct 1837

ctatagataa tatatgtgct gtgtgtaatt ttaaatttct ccaaccgtgc tgtacaaatg 1897

- 10 -

tgtggatttg taatcaggct attttgttgt tgttgttgtt gttcagagcc attaataataa 1957  
tatttaaagt tgagttcact ggataagttt ttcattctgc ccaaccattt ctaactgcc 2017  
aattgaattc aagaaaccga tgtgggtttt gtttcctgta caattatgag atataattct 2077  
ttttccatt gtaggtcttt tacaaaacaa gaaaataatt tatttttttg ttggtggata 2137  
aagaagtcaa gtatctgata ctttttattt acaaagtgtg atggttttgt atagtaggtt 2197  
ccaccctgag tattcctaaa agaaaaaaaa aaaaaaagct taaaaactct aacttcattct 2257  
gtgtttgtct tacgttgtct taatcgttgt acttacctta aaataaaccc atgttgtttt 2317  
ttctgcccaa agtttgaca gtgtgttgtt gttgttgcatt tttttacaaa cgaggtgtgt 2377  
ttgcaaacc acctgctttg attatttttg ttacacaggt gggatatatgt gtagacacat 2437  
aaaaacgacc agagaatagg agcacacacc tgctgtcttg tttagtgaac gaaaaaggt 2497  
tttgattaat tttaaaatcc cactctagga tttttctttt tcgagaaacc gccagttgg 2557  
agggggctgc ctgaaggacc ggaccatgag ttgccgtga tgcattttct taaatgcaca 2617  
aaaacatgct aattgtcaa acaaacagt ccaactccat tcagtgtcca gccgtcccca 2677  
gtttaggagg tgaaggaagg gaagaataaa catttccgt ttgctaactg caaccagggt 2737  
tgagtcctgc tttccccga ttttataaaa ttgagcctc ttgctgtgt ttaatagttt 2797  
tccagagaat ttgaactggg ccaatgaagg tctgaagggg acggattttc tagcgtttga 2857  
tatccatccc ccttagcggc cagatcagag ggaatttca gactttatta cttctcaatg 2917  
tcattgtctaa atctacacc tcattgcagt gaaaaattt aaaacctcat tacccttcaa 2977  
aaataattta tgatattttt agagttctaa attcaagttt ttcaatatgt taaataatag 3037  
agattatttt ttgttttcaa tgttaatatc tcgtctttta catttttaat agtaacatag 3097  
ttttgtgaa atgtagctga cgaaatggct ttattatcta tttcaatggc tgaagtccac 3157  
cactcccctg ctggcctcta tgtgtgaatt tggggaccaa agcttcatca attcccaccc 3217  
cagcaggtga gctgtacctt gctaattgtg aagttctttg tgagcttaac gtttcaagac 3277  
cagatgattt tgctaaagg gattttgctt gatgcagtg cgctgaacgt aaccgggtg 3337  
ttttgtcgt gttgttttca acatggcact ttatctccac gctatgttga aatagaatta 3397  
ggggaagctt aaagcataat aattgtccc acatgtgcaa cacagactct ttcaatctgt 3457  
ggccccagag gtggcacaca gttaagactt ggcggctgtc tcattctttt tcataatgtg 3517  
cggttcccg ggtgtccggg tgctagactt tcagcaggcc ccaggccaga cgggttttg 3577  
ttgagtgaac aggaggagga agttaaggag gtaggggtg ggagagacc tctccaagct 3637  
gcagaagaag gtggccaag ctccttgctt gcgtctgccc tgatggttt attttacttc 3697

- 11 -

tgctcgcttc atgctatttg ccccaggaga agaggagagt attccagacg gtaagcgagc 3757  
 tggctttttc ccttccttag acgttttaaa gaaatctttc tgaaagcttg ccctcatcgt 3817  
 aagctttgaa accgttggtg tcctgttagt ggcgagggct gagagacacg cggagaaata 3877  
 aaggagagcg acggtgtggc tgagagcccc caggtctgct gttgaaacta agctgggctt 3937  
 ttgcaccttt aggaagcctt tttaaagaag tcctgctgtg tgggggcccgg aagcccaagt 3997  
 gagtgggcct tgtggaggtt atcgggaggg gtctttacca ctcttgggg aacgtgggca 4057  
 acgggggggat tgtatctgaa gctttattca ggtcttcggc ggcagcagag tggagaacca 4117  
 ggcccttagt gtgtagcggc ctggggattt tgggactcat c 4158

&lt;210&gt; 4

&lt;211&gt; 417

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

Met	Asn	Ser	Gln	Glu	Gly	Cys	Lys	Glu	Thr	Leu	Asn	Gly	Ile	Tyr	Gln
1				5					10					15	
Phe	Ile	Met	Asp	Arg	Phe	Pro	Phe	Tyr	Arg	Glu	Asn	Lys	Gln	Gly	Trp
			20					25					30		
Gln	Asn	Ser	Ile	Arg	His	Asn	Leu	Ser	Leu	Asn	Glu	Cys	Phe	Val	Lys
	35						40					45			
Val	Pro	Arg	Asp	Asp	Lys	Lys	Pro	Gly	Lys	Gly	Ser	Tyr	Trp	Thr	Leu
	50					55					60				
Asp	Pro	Asp	Ser	Tyr	Asn	Met	Phe	Glu	Asn	Gly	Ser	Phe	Leu	Arg	Arg
65					70					75				80	
Arg	Arg	Arg	Phe	Lys	Lys	Lys	Asp	Val	Ser	Lys	Glu	Lys	Glu	Glu	Arg
				85					90					95	
Ala	His	Leu	Lys	Glu	Pro	Pro	Pro	Ala	Ala	Ser	Lys	Gly	Ala	Pro	Ala
			100					105					110		
Thr	Pro	His	Leu	Ala	Asp	Ala	Pro	Lys	Glu	Ala	Glu	Lys	Lys	Val	Val
		115					120					125			
Ile	Lys	Ser	Glu	Ala	Ala	Ser	Pro	Ala	Leu	Pro	Val	Ile	Thr	Lys	Val
	130						135				140				
Glu	Thr	Leu	Ser	Pro	Glu	Ser	Ala	Leu	Gln	Gly	Ser	Pro	Arg	Ser	Ala
145					150					155				160	
Ala	Ser	Thr	Pro	Ala	Gly	Ser	Pro	Asp	Gly	Ser	Leu	Pro	Glu	His	His
				165					170					175	
Ala	Ala	Ala	Pro	Asn	Gly	Leu	Pro	Gly	Phe	Ser	Val	Glu	Asn	Ile	Met
			180					185					190		

- 12 -

```

Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser Pro Gly Ala Gly
    195                                200                                205

Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro Tyr Ala Ala Ala
    210                                215                                220

Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly Leu Glu Ala Gly
    225                                230                                235                                240

Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met Ser Leu Tyr Thr
    245                                250                                255

Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro Ala Leu Asp Glu
    260                                265                                270

Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro Leu Ser Ala Leu
    275                                280                                285

Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala Thr Gly His His
    290                                295                                300

His Gln His His Gly His His His Pro Gln Ala Pro Pro Pro Pro Pro
    305                                310                                315                                320

Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala Ala Ala Ala Gln
    325                                330                                335

Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu Asn His Leu Pro
    340                                345                                350

Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro Asn Val Arg Glu
    355                                360                                365

Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser Thr Leu Gly Glu
    370                                375                                380

Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro Tyr Arg Ser Thr
    385                                390                                395                                400

Pro Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr Asp Cys Thr Lys
    405                                410                                415

```

Tyr

```

<210> 5
<211> 6021
<212> DNA
<213> Mus musculus

<220>
<221> exon
<222> (1649) .. (4348)

<300>
<308> GenBank/Y08222

```

- 13 -

&lt;309&gt; 1997-05-14

&lt;300&gt;

&lt;301&gt; Miura, N

&lt;303&gt; Genomics

&lt;304&gt; 41

&lt;306&gt; 489-492

&lt;307&gt; 1997

&lt;400&gt; 5

ctcagagtcaa	aggtagcaca	cataaaacct	atthttgctgc	ttcggtagct	caagcaatgc	60
cactaaagt	tcctcaccg	ccaaagctga	aacagtgcgt	tctaattctt	caaagccttt	120
tgccgaaaat	ctaaagggg	tgggggggcta	tgggtggtggc	gtggggggggg	ggtcggagaa	180
gaagaaagac	tgagacaaat	gtttttatctg	tcgccttctt	ccctacccaa	cgggaccaac	240
aacttcaga	aggttctgcg	aggcatagag	ccattccgta	gggacatctc	ggtgcttctg	300
aggaagcggg	ccgagcagg	atccgatgac	gactggagat	gttgaaggaa	taaataccag	360
tccacaaata	aacaaactgt	ccccgggatt	cctagaggga	aggagcacgc	ttgaaggtcg	420
gggaactccg	agtcgctgtg	cgtaagggtt	ggcataaaat	taaaaaaaaa	aaaagtcctt	480
cagttaccag	gcccctctaag	gagccccctg	tcctcagctc	accttatcaa	aactcagtaa	540
aacaaacagc	ctgaaataca	gtcaattttac	aggatcccaa	agatgctgac	cgcggagtgg	600
gaccacgcc	gggccccggc	aacagctagg	gaagcgggtc	cgaggctaca	cagtgcgcgc	660
ctcctttgcg	tttccagtga	cgaagccggc	gatggagtgc	aggcttggag	ctccccacgc	720
cgaacgggga	caccagctcc	cgggggctgg	ctgccttgct	ctaaccctca	gacagcgctt	780
tcataggtgg	gagaaaggga	gaggccggga	tggatggcag	ggaaagctag	ccctcgtcta	840
tgccggagag	gagaccagga	aagcaacagt	tgggttcacg	cgcttccctg	aacccacga	900
aattgtttgg	aggactcaga	tggatcacct	aagtagcagc	gaagacgaag	gaccaatggt	960
tccttaggtg	ttaccttccc	agtttggcat	tcccactaag	ccttccctcc	cagcccgacc	1020
ccgtcgtgaa	ggggagagga	accgaattct	ccaacccggc	ctcctttgtg	ggctcttcct	1080
caacctggaa	gcgtcctgtg	aattatccat	cactgcattc	aacaggccct	acacgctcag	1140
tccgtttgct	ctgaacccat	tacaactagg	ccccgataat	taagaaatct	aattattcgc	1200
ctcttcaccc	attaataata	ataaaaaaaaa	aatctccagg	ctctttccta	cttacaaggt	1260
cttgggggca	aatctctgcc	caacttcac	aattcgatgt	tatatattca	actaaacttc	1320
tttttatttt	ccaaaggaac	agggttttta	atthttgctc	tggacacgtg	gtctcgtaa	1380
acaaaatgtg	ataataaaat	aaaattttat	aagatgtaac	tcatttttaa	aagtcctcaa	1440
gttaacttga	gctggggggg	ggggagatct	ggctaagagc	atctgggtct	tagagccgac	1500
ggattcaggc	gctcctcgtt	ttgattggtg	ccatccttct	cgcagctgcc	agatgattgg	1560
tgcaaaactt	ctggaggggg	cgcggcctga	agaaagtaaa	aactcgcttt	gagccagaag	1620
acttttgaaa	cttttcccaa	tccctaaaag	ggactttgct	tctttttccg	ggctcggcc	1680
cgtcccttcc	tcggacccta	gctcgtgac	gctgcgggct	gcagttctcc	tggcggggcc	1740
cgagagccgc	tgtctccttt	tctagcactc	ggaagggtcg	gtgtcgctcc	acggtcgcgc	1800
gtggcgctcg	tgccgccagc	tcagggtctg	cacccgccaa	gccgagagt	cgcggccagc	1860
ggggccgcct	gccgtgcacc	cttcaggatg	ccgatccgcc	cggtcggtcg	aacccgagcg	1920
ccggcgctct	ccgcgcgtgg	accgcgaggc	tgccccgagt	cggggctgcc	tgcatcgctc	1980
cgtcccttcc	tgcctcctcg	ctccgggctt	cgctcgccgc	gggcccgcagt	cgggtgcgcg	2040
aggcggcgac	cgggcgtctg	ggagcgagca	tgcaggcgcg	ttactcggtg	tcggacccca	2100
acgccttggt	agtggtaccc	tatttgagt	agcaaaacta	ctaccgggag	gccggcagct	2160
acggcgccat	ggccagcccc	atgggcgtct	actccggcca	cccggagcag	tacggcgccg	2220
gcattggccg	ctcctacgcg	ccctaccacc	accagccgcg	ggcgcccaag	gacctggtga	2280
agccgcccta	cagctatata	gcgctcatca	ccatggcgat	ccagaacgcg	ccagagaaga	2340
agatcactct	gaacggcatc	taccagttca	tcattggaccg	tttcccttcc	taccgcgaga	2400
acaagcaggg	ctggcagaac	agcatccgcc	acaacctgtc	actcaatgag	tgtctcgtga	2460
aagtgcgcgc	cgacgacaag	aagccgggca	agggcagcta	ctggacgctc	gacccggact	2520
cctacaacat	gttcgagaat	ggcagcttcc	tgcggcgggc	gcggcgcttc	aagaagaagg	2580
atgtgcccaa	ggacaaggag	gagcggggcc	acctcaagga	gccgcccctg	accacggcca	2640
aggcgctccc	gacagggacc	ccggtagctg	acggggccaa	ggaggccgag	aagaaagtcg	2700
tggttaagag	cgaggcgggc	tccccgcgc	tgccggtcat	caccaagggtg	gagacgctga	2760
gccccgaggg	agcgtgcag	gccagtccgc	gcagcgcatc	ctccacgccc	gcaggttccc	2820
cagacggctc	gctgcgggag	caccacgcgc	cggcgcccta	cgggctgccc	ggcttcagcg	2880
tggagaccat	catgacgctg	cgcacgtcgc	ctccggggcg	cgatctgagc	ccagcgggcg	2940



```

cgcgcgcgcg cctggtggtg ccaccgctgg cactgccata cgccgcagcg ccaccgccc 3000
cttacacgca gccgtgcgcg cagggcctgg aggtgcggg ctcgcgcggc taccagtga 3060
gtatgcgggc tatgagtctg tacaccgggg ccgagcggcc cgcgcacgtg tgcgttccgc 3120
ccgcgctgga cgaggtctctg tcggaccacc cgagcggccc cggctccccg ctcggcgccc 3180
toaacctcgc agcgggtcag gaggggcgcg tgggggcctc gggtcaccac caccagcatc 3240
acggccacct ccaccgcgag gcgccaccgc ccgcccgcga gcccctccc gcgcgcgagc 3300
ccgccaccca gccacacctc tggatatga accacggcgg ggacctgagc cacctccccg 3360
gccacacggt tgcaacccaa cagcaaaact tcccacacgt ccgggagatg ttcaactcgc 3420
accggctagg actggacaac tcgtccctcg gggagtccca ggtgagcaat gcgagctgtc 3480
agctgcctta tcgagctacg ccgtccctct accgccacgc agccccctac tcttacgact 3540
gcaccaaata ctgaggctgt ccagtccgct ccagccccag gaccgcaccg gcttcgcctc 3600
ctccatggga acctctctcg acggagccgc agaaagcgac ggaaagcgcc cctctctcag 3660
aaccaggagc agagagctcc gtgcaactcg caggtaaact atccgcagct cagtttgaga 3720
tctcagcgag tccctctaag ggggatgcag cccagcaaaa cgaaatacag attttttttt 3780
taattccttc ccctaccag atgtgcgcg tgctcccttg gggcttcata gattagctta 3840
tggaccaaac ccatagggac ccctaatagac ttctgtggag attctccacg ggcgcaagag 3900
gtctctcccg ataagggtgcc ttctgtaaac gagtgcggat ttgtaaccag gctattttgt 3960
tcttgcccag agcctttaat ataataatga aagtgtgtc cactggataa ggtttcgtct 4020
tgcccaactg ttactgccaa attgaattca agaaacgtgt gtgggtcttt tctccccacg 4080
tcaccatgat aaaaataggtc cctcccaaaa ctgtaggtct ttacaaaaac aagaaaaata 4140
tttatttttt tgtgtgtgtt ggataacgaa attaatgata ggatactttt aatttaggaa 4200
gtgcatggct ttgtacagta gatgccatct ggggtattcc aaaaacacac caaaagactt 4260
taaaatttca atctcacctg tgtttgtctt atgtgatctc agtgttgtat ttaccttaaa 4320
ataaacccgt gttgtttttc tgccaaaagt tcggacagag tctttgtgtt cttgaatttt 4380
aaaagggaaa ttgtagtaag ccagttgtga ttgatttttg tgatgcaggt tggcctggta 4440
acgtggatgc atatacaggt tacaggacga tggagctctc gattagtaat agaaggggct 4500
cttgatttgt tgaactatcc cgtcctgaga tatttttgtt ttctgctcga ggtaatctga 4560
gaaactgttc tccatccaca cagcgacagg gctgcctgag ggcaacgtcc tgctggcctg 4620
ttaacgaaat gctttgcggg atgcagaaaa ctgttgccaa ttgtcaaaac aaaatggtgt 4680
caccctgtct cggtgtccag ctgtcctctg ttagagggga gaaaccgaga aaggacaaac 4740
ggcctgcagc ttgctaacct cagcgtagca ggagcctggg tgagtgtctc gctccctcca 4800
tttccttaga tgcggacttg ttgcccctgt tggcgtttta agagtgccag caagaagcaa 4860
agaggggttg taggtctctg gtatttaact gccggctttg ggatcagatt agaagtgaat 4920
ttcagtctga ttattttctt aatttgggct ttaaatattt tactccggcg tggtggaata 4980
agaagccact gtgcgcctcc agcatgatat ttagcgctg aaatggctct ggttttcagc 5040
atgctaagta acaggagatt atttttcttt tgattcttgt atttcatttc tttaaaaaaa 5100
aaaaaggaaa tagatcggga caaactctct aaaaatgtacc tggctggctg ggggtggggtc 5160
cttaccatc tgctgcctga aagatacagc ttcagcacag gcctgcgtgt tggactttag 5220
gcataatcag gatccacg ccagttggta acctggactg tgctaataga agttctttct 5280
gcacagaaca tgtaggccag gaggaggcag ggacccggga ggggggtgga ctttgaggt 5340
catctgctta gcttagtggt ggccacgggt taacacgtat atagtgttac tgtttgaaac 5400
tocaagtttt atatctgtgc tgttttgatg tagaatttg ggaggttct gatgatacta 5460
ccctaccggt gtatgtaaga cagtctttca acctgcagtg ccagaatgtg acccacactt 5520
cagtatcttc cataaagtgg ggggactaag aactggacag ggtgctgtg gaggggggca 5580
ggccagggtg atcttgggtc ctgagcagag cagagagctt aggaaggggt cgggagatct 5640
ctgggttctc ccaacactgg ttcatctttg catggctctc ttcaaacctc ttgcccagg 5700
agaagcgagc tttgtccaag ccagctggct cgctcctttc ccagatgttt taggggctc 5760
cctgaaagct tgccctcctc ttaagattca gaactcctga cccagggaaa gataggaggc 5820
tttgtggat ggagcttttt tttaaagagg accgtctctg ttctcaagta ggtagctaga 5880
gagaagcccc ctggagcagg cctacttgt gactgtcagg gaaccaggt tgtgtgtag 5940
gcttttccca ggccctccag agcagcgggt tgaaaaaatg cggtcctggg aaaagtgggt 6000
ctggggtgtt gcttctcga g 6021

```

&lt;210&gt; 6

&lt;211&gt; 2712

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

- 15 -

<220>  
 <221> CDS  
 <222> (422)..(1906)

<300>  
 <308> GenBank/Y08222  
 <309> 1997-05-14

<300>  
 <301> Miura, N  
 <303> Genomics  
 <304> 41  
 <306> 489-492  
 <307> 1997

<400> 6  
 agggactttg cttctttttc cgggctcggc cgcgagcct ctccggaccc tagctcgctg 60  
 acgtgcggg ctgcagttct cctggcgggg cccgagagcc gctgtctcct tttctagcac 120  
 tcggaagggc tgggtgctgct ccacggtcgc gcgtggcgtc tgtgccgcca gctcagggct 180  
 gccacccgcc aagccgagag tgcgcggcca gcggggccgc ctgccgtgca cccttcagga 240  
 tgccgatccg cccggtcggc tgaacccgag cgccggcgtc ttccgcgcgt ggaccgcgag 300  
 gctgccccga gtcggggctg cctgcacgc tccgtccctt cctgctctcc tgctccgggc 360  
 ctgctcgcgc gcggggccga gtcggtgcgc gcaggcggcg accggggcgtc tgggacgcag 420  
 c atg cag gcg cgt tac tgc gta tgc gac ccc aac gcc ctg gga gtg gta 469  
 Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala Leu Gly Val Val  
 1 5 10 15  
 ccc tat ttg agt gag caa aac tac tac cgg gcg gcc ggc agc tac ggc 517  
 Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala Gly Ser Tyr Gly  
 20 25 30  
 ggc atg gcc agc ccc atg ggc gtc tac tcc ggc cac ccg gag cag tac 565  
 Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His Pro Glu Gln Tyr  
 35 40 45  
 ggc gcc ggc atg ggc cgc tcc tac gcg ccc tac cac cac cag ccc gcg 613  
 Gly Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His His Gln Pro Ala  
 50 55 60  
 gcg ccc aag gac ctg gtg aag ccg ccc tac agc tat ata gcg ctc atc 661  
 Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser Tyr Ile Ala Leu Ile  
 65 70 75 80  
 acc atg gcg atc cag aac gcg cca gag aag aag atc act ctg aac ggc 709  
 Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys Ile Thr Leu Asn Gly  
 85 90 95  
 atc tac cag ttc atc atg gac cgt ttc ccc ttc tac cgc gag aac aag 757  
 Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn Lys  
 100 105 110

- 16 -

cag ggc tgg cag aac agc atc cgc cac aac ctg tca ctc aat gag tgc	805
Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu Cys	
115 120 125	
ttc gtg aaa gtg ccg cgc gac gac aag aag ccg ggc aag ggc agc tac	853
Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser Tyr	
130 135 140	
tgg acg ctc gac ccg gac tcc tac aac atg ttc gag aat ggc agc ttc	901
Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser Phe	
145 150 155 160	
ctg cgg cgg cgg cgg cgc ttc aag aag aag gat gtg ccc aag gac aag	949
Leu Arg Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Pro Lys Asp Lys	
165 170 175	
gag gag cgg gcc cac ctc aag gag ccg ccc tcg acc acg gcc aag ggc	997
Glu Glu Arg Ala His Leu Lys Glu Pro Pro Ser Thr Thr Ala Lys Gly	
180 185 190	
gct ccg aca ggg acc ccg gta gct gac ggg ccc aag gag gcc gag aag	1045
Ala Pro Thr Gly Thr Pro Val Ala Asp Gly Pro Lys Glu Ala Glu Lys	
195 200 205	
aaa gtc gtg gtt aag agc gag gcg gcg tcc ccc gcg ctg ccg gtc atc	1093
Lys Val Val Val Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val Ile	
210 215 220	
acc aag gtg gag acg ctg agc ccc gag gga gcg ctg cag gcc agt ccg	1141
Thr Lys Val Glu Thr Leu Ser Pro Glu Gly Ala Leu Gln Ala Ser Pro	
225 230 235 240	
cgc agc gca tcc tcc acg ccc gca ggt tcc cca gac ggc tcg ctg ccg	1189
Arg Ser Ala Ser Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu Pro	
245 250 255	
gag cac cac gcc gcg gcg cct aac ggg ctg ccc ggc ttc agc gtg gag	1237
Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val Glu	
260 265 270	
acc atc atg acg ctg cgc acg tcg cct ccg ggc ggc gat ctg agc cca	1285
Thr Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Asp Leu Ser Pro	
275 280 285	
gcg gcc gcg cgc gcc ggc ctg gtg gtg cca ccg ctg gca ctg cca tac	1333
Ala Ala Ala Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro Tyr	
290 295 300	
gcc gca gcg cca ccc gcc gct tac acg cag ccg tgc gcg cag ggc ctg	1381
Ala Ala Ala Pro Pro Ala Ala Tyr Thr Gln Pro Cys Ala Gln Gly Leu	
305 310 315 320	
gag gct gcg ggc tcc gcg ggc tac cag tgc agt atg ccg gct atg agt	1429
Glu Ala Ala Gly Ser Ala Gly Tyr Gln Cys Ser Met Arg Ala Met Ser	
325 330 335	
ctg tac acc ggg gcc gag ccg ccc gcg cac gtg tgc gtt ccg ccc gcg	1477
Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Val Cys Val Pro Pro Ala	
340 345 350	

- 17 -

ctg gac gag gct ctg tgc gac cac ccg agc ggc ccc ggc tcc ccg ctc 1525  
 Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Gly Ser Pro Leu  
 355 360 365

ggc gcc ctc aac ctc gca gcg ggt cag gag ggc gcg ttg ggg gcc tgc 1573  
 Gly Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Gly Ala Ser  
 370 375 380

ggt cac cac cac cag cat cac ggc cac ctc cac ccg cag gcg cca ccg 1621  
 Gly His His His Gln His His Gly His Leu His Pro Gln Ala Pro Pro  
 385 390 395 400

ccc gcc ccg cag ccc cct ccc gcg ccg cag ccc gcc acc cag gcc acc 1669  
 Pro Ala Pro Gln Pro Pro Pro Ala Pro Gln Pro Ala Thr Gln Ala Thr  
 405 410 415

tcc tgg tat ctg aac cac ggc ggg gac ctg agc cac ctc ccc ggc cac 1717  
 Ser Trp Tyr Leu Asn His Gly Gly Asp Leu Ser His Leu Pro Gly His  
 420 425 430

acg ttt gca acc caa cag caa act ttc ccc aac gtc cgg gag atg ttc 1765  
 Thr Phe Ala Thr Gln Gln Gln Thr Phe Pro Asn Val Arg Glu Met Phe  
 435 440 445

aac tgc cac cgg cta gga ctg gac aac tgc tcc ctc ggg gag tcc cag 1813  
 Asn Ser His Arg Leu Gly Leu Asp Asn Ser Ser Leu Gly Glu Ser Gln  
 450 455 460

gtg agc aat gcg agc tgt cag ctg ccc tat cga gct acg ccg tcc ctc 1861  
 Val Ser Asn Ala Ser Cys Gln Leu Pro Tyr Arg Ala Thr Pro Ser Leu  
 465 470 475 480

tac cgc cac gca gcc ccc tac tct tac gac tgc acc aaa tac tga 1906  
 Tyr Arg His Ala Ala Pro Tyr Ser Tyr Asp Cys Thr Lys Tyr  
 485 490 495

ggctgtccag tccgctccag cccagagacc gcaccggcct cgcctcctcc atgggaacct 1966  
 tcttcgacgg agccgcagaa agcgacggaa agcgcccctc tctcagaacc aggagcagag 2026  
 agctccgtgc aactgcagg taacttatcc gcagctcagt ttgagatctc agcgagtccc 2086  
 tctaaggggg atgcagccca gcaaaacgaa atacagattt tttttttaat tccttcccct 2146  
 acccagatgc tgcgcctgct cccttggggc ttcatagatt agcttatgga ccaaaccat 2206  
 agggaccctt aatgacttct gtggagattc tccacgggag caagaggtct ctccggataa 2266  
 ggtgccttct gtaaacgagt gcggatttgt aaccaggcta ttttgttctt gccagagacc 2326  
 tttaatataa tatttaaagt tgtgtccact ggataagggt tcgtcttgcc caactgttac 2386  
 tgccaaattg aattcaagaa acgtgtgtgg gtcttttctc cccacgtcac catgataaaa 2446  
 taggtccctc cccaaactgt aggtctttta caaaacaaga aaataattta tttttttgtt 2506  
 gttgttgat aacgaaatta agtatcggat acttttaatt taggaagtgc atggctttgt 2566

- 18 -

acagtagatg ccatctgggg tattccaaaa acacacacaaa agactttataa atttcaatct 2626  
cacctgtggtt tgtcttatgt gatctcagtg ttgtatttac cttaaaataa acccgtgttg 2686  
tttttctgcc caaaaaaaaa aaaaaa 2712

<210> 7  
<211> 494  
<212> PRT  
<213> Mus musculus

<400> 7  
Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala Leu Gly Val Val  
1 5 10 15  
Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala Gly Ser Tyr Gly  
20 25 30  
Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His Pro Glu Gln Tyr  
35 40 45  
Gly Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His His Gln Pro Ala  
50 55 60  
Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser Tyr Ile Ala Leu Ile  
65 70 75 80  
Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys Ile Thr Leu Asn Gly  
85 90 95  
Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn Lys  
100 105 110  
Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu Cys  
115 120 125  
Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser Tyr  
130 135 140  
Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser Phe  
145 150 155 160  
Leu Arg Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Pro Lys Asp Lys  
165 170 175  
Glu Glu Arg Ala His Leu Lys Glu Pro Pro Ser Thr Thr Ala Lys Gly  
180 185 190  
Ala Pro Thr Gly Thr Pro Val Ala Asp Gly Pro Lys Glu Ala Glu Lys  
195 200 205  
Lys Val Val Val Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val Ile  
210 215 220  
Thr Lys Val Glu Thr Leu Ser Pro Glu Gly Ala Leu Gln Ala Ser Pro  
225 230 235 240  
Arg Ser Ala Ser Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu Pro  
245 250 255  
Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val Glu  
260 265 270  
Thr Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Asp Leu Ser Pro  
275 280 285  
Ala Ala Ala Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro Tyr  
290 295 300  
Ala Ala Ala Pro Pro Ala Ala Tyr Thr Gln Pro Cys Ala Gln Gly Leu  
305 310 315 320  
Glu Ala Ala Gly Ser Ala Gly Tyr Gln Cys Ser Met Arg Ala Met Ser  
325 330 335  
Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Val Cys Val Pro Pro Ala  
340 345 350  
Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Gly Ser Pro Leu  
355 360 365

- 19 -

Gly Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Gly Ala Ser  
 370 375 380  
 Gly His His His Gln His His Gly His Leu His Pro Gln Ala Pro Pro  
 385 390 395 400  
 Pro Ala Pro Gln Pro Pro Ala Pro Gln Pro Ala Thr Gln Ala Thr  
 405 410 415  
 Ser Trp Tyr Leu Asn His Gly Gly Asp Leu Ser His Leu Pro Gly His  
 420 425 430  
 Thr Phe Ala Thr Gln Gln Gln Thr Phe Pro Asn Val Arg Glu Met Phe  
 435 440 445  
 Asn Ser His Arg Leu Gly Leu Asp Asn Ser Ser Leu Gly Glu Ser Gln  
 450 455 460  
 Val Ser Asn Ala Ser Cys Gln Leu Pro Tyr Arg Ala Thr Pro Ser Leu  
 465 470 475 480  
 Tyr Arg His Ala Ala Pro Tyr Ser Tyr Asp Cys Thr Lys Tyr  
 485 490

<210> 8  
 <211> 3289  
 <212> DNA  
 <213> Homo sapiens

<300>  
 <301> Miura, N  
 <303> Genomics  
 <304> 41  
 <306> 489-492  
 <307> 1997

<300>  
 <308> GenBank/Y08223  
 <309> 1997-05-14

<400> 8  
 gaattcggag gattaagttg tcagtcagca cgttgetacc ttccctctta tgcactccgc 60  
 tgcctggctc ctccggcggg agcgagggaa actcagtttg tagggtttac ctctaaaacc 120  
 tcgatagggt atccttgacg accccgagcc tggaaactcc ctgttgatga ttaattattt 180  
 gattaaataa gtataacatc caggagaggc cctgccattc caatccagcg cgtttgcttt 240  
 tgaatccatt acacctgggc ccccataatt aggaaatcta attattcgct tcatcactca 300  
 ttaataagaa aaatgtccca ggatcattgc tacttacaag gtctttggga gagatatttt 360  
 actctattaa tccattctat tttatatttc aaattgattt tttttaacag aggaaagtgg 420  
 ctatcttttt gttttgggca tgtgggcca ttacacaaaa tgtgatcata aaataaattt 480  
 taataagata taacttttta aaaagttttc aagtgaagac ggagtcgccg cggaggcccg 540  
 ggcggcgggg tcttagagcc gacggattcc tgcgctcctc gccccgattg gcgcggact 600  
 cctctcagct gccgggtgat tggctcaaag ttccgggagg ggcggtggcc cgaggaaaagt 660  
 aaaaactcgc tttcagcaag aagacttttg aaacttttcc caatccctaa aagggaacttg 720  
 gcctcttttt ctgggctcag cggggcagcc gctcggaccc cggcgcgctg accctcgggg 780  
 ctgccgattc gctgggggct tggagagcct cctgcgcccc tcctcgcgcg gcccgagggt 840  
 ccaccttggg ccccgaggcc cggcgtctcc gctgggtccg cggccgcccc cctgcccgcg 900  
 ctgcgcggcg cgggtcctgg agccagcgag gagcggggcc ggcgctgcgc ttgccgggg 960  
 cgcgccctcc aggatgccga tccgccggt ccgctgaaaag cgcgcgcccc tgctcggccc 1020  
 gagcgacgac gaccgcgcac cctcgccccg gaggtgccca ggagaccggg gccgccctc 1080  
 ccgetcccc cctctcccc tctggtctc tcgctctcag ggccccctc 1140  
 gctcccccg ccgcagtcg tgcgcgagg cgccggcgag ccgtctcgga agcagcatgc 1200  
 aggcgcgcta ctccgtgtcc gaccccaacg ccctgggagt ggtgccctac ctgagcgagc 1260  
 agaattacta ccgggctgcg ggcagctacg cgggcattgc cagccccatg ggcgtctatt 1320  
 ccggccaccc ggagcagtac agcgcgggga tgggccgctc ctacgcgccc taccaccacc 1380

- 20 -

```

accagccccg ggcgccctaag gacctggtga agccgcccta cagctacatc gcgctcatca 1440
ccatggccat ccagaacgcg cccgagaaga agatcacctt gaacggcatc taccagttca 1500
tcatggaccg cttccccctt taccgggaga acaagcaggg ctggcagaac agcatccgcc 1560
acaacctctc gctcaacgag tgcttcgtca aggtgccccg cgacgacaag aagcccgga 1620
agggcagtta ctggaccctg gacccggact cctacaacat gttcgagaac ggcagcttcc 1680
tgccggcgccg gcgcgcttc aaaaagaagg acgtgtccaa ggagaaggag gagcgggccc 1740
acctcaagga gccgcccccg gcggcggtcca agggcgcccc gccaccccc cacctagcgg 1800
acgcccccaa ggaggccgag aagaagggtg tgatcaagag cgaggcgcg tccccggcgc 1860
tgccggctcat caccaagggt gagacgctga gccccgagag cgcgctgcag ggcagccgc 1920
gcagcgcggc ctccacgccc gccggctccc ccgacggttc gctgccggag caccacgccc 1980
cggcgcccaa cgggctgcct ggcttcagcg tggagaacat catgaccctg cgaacgtcgc 2040
cgccggcgcg agagctgagc ccggggggcg gacgcgcggg cctggtggtg ccgcccgtgg 2100
cgctgccata cgccgcgcgc ccgcccgcgc cctacggcca gccgtgcgct cagggcctgg 2160
aggccggggc cgccgggggc taccagtgc gcatgcgagc gatgagcctg tacaccgggg 2220
ccgagcgggc ggcgcacatg tgcgtccgc ccgccctgga cgaggccctc tcggaccacc 2280
cgagcgggcc cactcgccc ctgagcgctc tcaacctgc cgccggccag gaggcgcg 2340
tcgcccgaac gggccaccac caccagcacc acggccacca ccaccgcag gcgcccgcgc 2400
ccccgcggc tccccagccc cagccgacgc cgacgcccgg gcccgccgcg gcgcaggcgg 2460
cctcctggta tctcaaccac agcggggacc tgaaccacct ccccgccac acgttcgcgg 2520
cccagcgaca aactttcccc aacgtgcggg agatgttcaa ctcccaccgg ctggggattg 2580
agaactcgac cctcggggag tcccaggtga gtggcaatgc cagctgccag ctgccctaca 2640
gatccacgcc gcctctctat cgccacgcag cccctactc ctacgactgc acgaaatact 2700
gacgtgtccc gggacctccc ctcccggccc cgctccggtc tcgcttccc gcccccaccc 2760
aaccagacaa ttaaggggt gcagagacgc aaaaaagaaa caaaacatgt ccaccaacct 2820
tttctcagac ccgggagcag agagcgggca cgctagcccc cagccgtctg tgaagagcgc 2880
aggtaacttt aattcgccgc ccggtttctg ggatcccagg aaaccctcc aaagggacgc 2940
agcccaacaa aatgagtatt ggtcttaaaa tccccctccc ctaccaggac ggctgtgctg 3000
tgctcgacct gagctttcaa aagttaagtt atggacccaa atcccatagc gagcccctag 3060
tgactttctg taggggtccc cataggtgta tgggggtctc tatagataat atatgtgctg 3120
tgtgtaattt taaatttctc caaccgtgct gtacaaatgt gtggatttgt aatcaggcta 3180
ttttgttgtt gttgttgttg ttcagagcca ttaataataat atttaagtt gagttcactg 3240
gataagtttt tcattctgcc caaccatttc taactgccaa attgaattc 3289

```

<210> 9  
 <211> 1506  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (1)..(1506)

<300>  
 <308> GenBank/NM\_005251  
 <309> 1999-12-23

<400> 9  
 atg cag gcg cgc tac tcc gtg tcc gac ccc aac gcc ctg gga gtg gtg 48  
 Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala Leu Gly Val Val  
 1 5 10 15  
  
 ccc tac ctg agc gag cag aat tac tac cgg gct gcg ggc agc tac ggc 96  
 Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala Gly Ser Tyr Gly  
 20 25 30  
  
 ggc atg gcc agc ccc atg gcc gtc tat tcc ggc cac ccg gag cag tac 144  
 Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His Pro Glu Gln Tyr  
 35 40 45

- 21 -

agc gcg ggg atg ggc cgc tcc tac gcg ccc tac cac cac cac cag ccc	192
Ser Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His His His Gln Pro	
50 55 60	
gcg gcg cct aag gac ctg gtg aag ccg ccc tac agc tac atc gcg ctc	240
Ala Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser Tyr Ile Ala Leu	
65 70 75 80	
atc acc atg gcc atc cag aac gcg ccc gag aag aag atc acc ttg aac	288
Ile Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys Ile Thr Leu Asn	
85 90 95	
ggc atc tac cag ttc atc atg gac cgc ttc ccc ttc tac cgg gag aac	336
Gly Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn	
100 105 110	
aag cag gcc tgg cag aac agc atc cgc cac aac ctc tcg ctc aac gag	384
Lys Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu	
115 120 125	
tgc ttc gtc aag gtg ccc cgc gac gac aag aag ccc ggc aag ggc agt	432
Cys Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser	
130 135 140	
tac tgg acc ctg gac ccg gac tcc tac aac atg ttc gag aac ggc agc	480
Tyr Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser	
145 150 155 160	
ttc ctg cgg cgc cgg cgg cgc ttc aaa aag aag gac gtg tcc aag gag	528
Phe Leu Arg Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Ser Lys Glu	
165 170 175	
aag gag gag cgg gcc cac ctc aag gag ccg ccc ccg gcg gcg tcc aag	576
Lys Glu Glu Arg Ala His Leu Lys Glu Pro Pro Pro Ala Ala Ser Lys	
180 185 190	
ggc gcc ccg gcc acc ccc cac cta gcg gac gcc ccc aag gag gcc gag	624
Gly Ala Pro Ala Thr Pro His Leu Ala Asp Ala Pro Lys Glu Ala Glu	
195 200 205	
aag aag gtg gtg atc aag agc gag gcg gcg tcc ccg gcg ctg ccg gtc	672
Lys Lys Val Val Ile Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val	
210 215 220	
atc acc aag gtg gag acg ctg agc ccc gag agc gcg ctg cag ggc agc	720
Ile Thr Lys Val Glu Thr Leu Ser Pro Glu Ser Ala Leu Gln Gly Ser	
225 230 235 240	
ccg cgc agc gcg gcc tcc acg ccc gcc ggc tcc ccc gac ggt tcg ctg	768
Pro Arg Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu	
245 250 255	
ccg gag cac cac gcc gcg gcg ccc aac ggg ctg cct ggc ttc agc gtg	816
Pro Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val	
260 265 270	



- 22 -

gag aac atc atg acc ctg cga acg tcg ccg ccg ggc gga gag ctg agc	864
Glu Asn Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser	
275 280 285	
ccg ggg gcc gga cgc gcg ggc ctg gtg gtg ccg ccg ctg gcg ctg cca	912
Pro Gly Ala Gly Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro	
290 295 300	
tac gcc gcc gcg ccg ccc gcc gcc tac ggc cag ccg tgc gct cag ggc	960
Tyr Ala Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly	
305 310 315 320	
ctg gag gcc ggg gcc gcc ggg ggc tac cag tgc agc atg cga gcg atg	1008
Leu Glu Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met	
325 330 335	
agc ctg tac acc ggg gcc gag cgg ccg gcg cac atg tgc gtc ccg ccc	1056
Ser Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro	
340 345 350	
gcc ctg gac gag gcc ctc tcg gac cac ccg agc ggc ccc acg tcg ccc	1104
Ala Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro	
355 360 365	
ctg agc gct ctc aac ctc gcc gcc ggc cag gag ggc gcg ctc gcc gcc	1152
Leu Ser Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala	
370 375 380	
acg ggc cac cac cac cag cac cac ggc cac cac cac ccg cag gcg ccg	1200
Thr Gly His His His Gln His His Gly His His His Pro Gln Ala Pro	
385 390 395 400	
ccg ccc ccg ccg gct ccc cag ccc cag ccg acg ccg cag ccc ggg gcc	1248
Pro Pro Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala	
405 410 415	
gcc gcg gcg cag gcg gcc tcc tgg tat ctc aac cac agc ggg gac ctg	1296
Ala Ala Ala Gln Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu	
420 425 430	
aac cac ctc ccc ggc cac acg ttc gcg gcc cag cag caa act ttc ccc	1344
Asn His Leu Pro Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro	
435 440 445	
aac gtg cgg gag atg ttc aac tcc cac ccg ctg ggg att gag aac tcg	1392
Asn Val Arg Glu Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser	
450 455 460	
acc ctc ggg gag tcc cag gtg agt ggc aat gcc agc tgc cag ctg ccc	1440
Thr Leu Gly Glu Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro	
465 470 475 480	
tac aga tcc acg ccg cct ctc tat cgc cac gca gcc ccc tac tcc tac	1488
Tyr Arg Ser Thr Pro Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr	
485 490 495	
gac tgc acg aaa tac tga	1506
Asp Cys Thr Lys Tyr	
500	

- 23 -

<210> 10  
 <211> 501  
 <212> PRT  
 <213> Homo sapiens

<400> 10  
 Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala Leu Gly Val Val  
 1 5 10 15  
 Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala Gly Ser Tyr Gly  
 20 25 30  
 Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His Pro Glu Gln Tyr  
 35 40 45  
 Ser Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His His His Gln Pro  
 50 55 60  
 Ala Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser Tyr Ile Ala Leu  
 65 70 75 80  
 Ile Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys Ile Thr Leu Asn  
 85 90 95  
 Gly Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn  
 100 105 110  
 Lys Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu  
 115 120 125  
 Cys Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser  
 130 135 140  
 Tyr Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser  
 145 150 155 160  
 Phe Leu Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Ser Lys Glu  
 165 170 175  
 Lys Glu Glu Arg Ala His Leu Lys Glu Pro Pro Pro Ala Ala Ser Lys  
 180 185 190  
 Gly Ala Pro Ala Thr Pro His Leu Ala Asp Ala Pro Lys Glu Ala Glu  
 195 200 205  
 Lys Lys Val Val Ile Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val  
 210 215 220  
 Ile Thr Lys Val Glu Thr Leu Ser Pro Glu Ser Ala Leu Gln Gly Ser  
 225 230 235 240  
 Pro Arg Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu  
 245 250 255  
 Pro Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val  
 260 265 270  
 Glu Asn Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser  
 275 280 285  
 Pro Gly Ala Gly Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro  
 290 295 300  
 Tyr Ala Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly  
 305 310 315 320  
 Leu Glu Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met  
 325 330 335  
 Ser Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro  
 340 345 350  
 Ala Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro  
 355 360 365  
 Leu Ser Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala  
 370 375 380  
 Thr Gly His His His Gln His His Gly His His Pro Gln Ala Pro  
 385 390 395 400

- 24 -

```

Pro Pro Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala
      405      410      415
Ala Ala Ala Gln Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu
      420      425      430
Asn His Leu Pro Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro
      435      440      445
Asn Val Arg Glu Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser
      450      455      460
Thr Leu Gly Glu Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro
      465      470      475      480
Tyr Arg Ser Thr Pro Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr
      485      490      495
Asp Cys Thr Lys Tyr
      500

```

```

<210> 11
<211> 327
<212> DNA
<213> Homo sapiens

```

```

<300>
<308> GenBank/AW271272
<309> 2000-01-03

```

```

<300>
<301> Strausberg, Robert
<303> Trends Genet.
<304> 16
<305> 3
<306> 103-106
<307> 2000

```

```

<400> 11
tttttttttac attttcgtct tctgtttottg tgattggaaa taagtggcac gccccattgc 60
cttctagtcg cctccccgaa gcgaagagggc cgaagcgaag aggcctgggtg ggttgtctca 120
acatcctttt gctgagaatc gaatacgcag ccgatgaaca gccaggaagg gtgcaaggaa 180
accttgaacg gcatctacca gttcatcatg gaccgcttcc ccttctaccg ggagaacaag 240
caggggctggc agaacagcat ccgccacaac ctctcgctca acgagtgcctt cgtcaagggtg 300
ccccgcgacg acaagaagcc cggcaag                                     327

```

```

<210> 12
<211> 147
<212> DNA
<213> Homo sapiens

```

```

<300>
<308> GenBank/AW793237
<309> 2000-05-16

```

```

<300>
<301> Dias Neto, E
<303> Proc. Natl. Acad. Sci. U.S.A.
<304> 97
<306> 3491-3496
<307> 2000

```

- 25 -

&lt;400&gt; 12

```
ccgtctgaga atcgaatacg cagccgatga acagccagga agggtgcaag gaaaccttga 60
acggcatcta ccagttcatc atggaccgct tcccccttcta ccgggagAAC aagcagggct 120
ggcagaacag catccgccac aacctct                                     147
```

00298-SE 2

Fig. 1

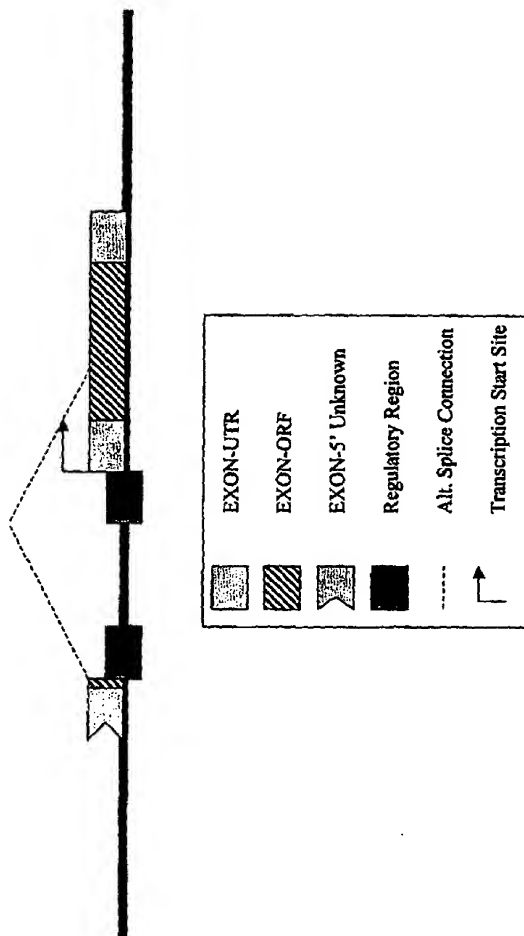


Fig. 2

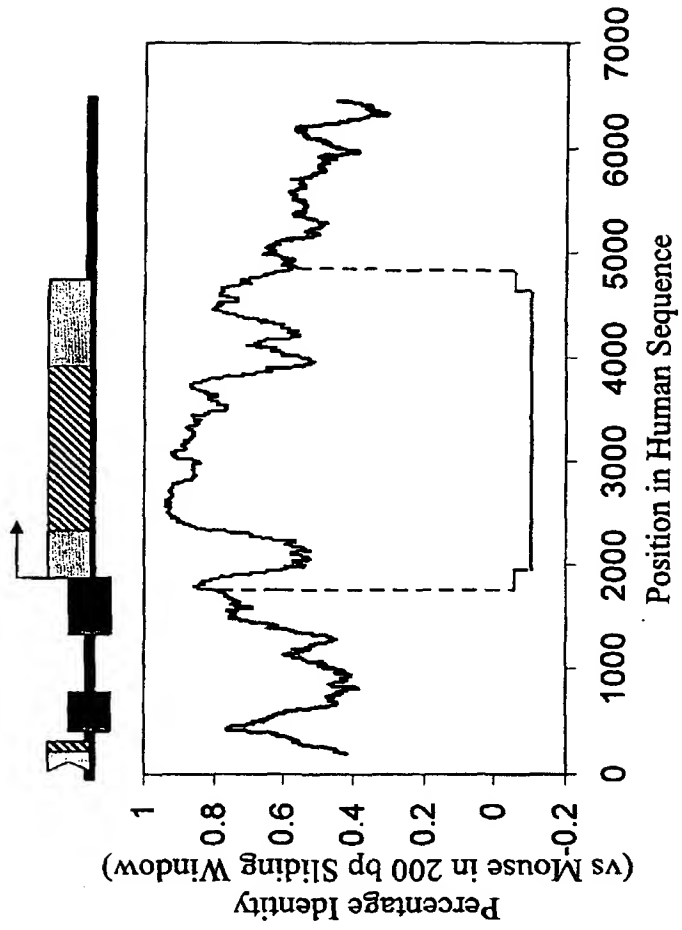


Fig. 3

	200	Splice	E-box-like
HUMAN		<u>AAGGGTGAAGGAAAC</u>	<u>CTGAAATACAAATGTTCTCCCTGAAGCCCTCTCCCTGCCCAAC</u>
MOUSE		<u>CGGAGAAGAAAGAAAG</u>	<u>CTGAG--ACAAATGTTTATCTGTGCGCTTCTTCCCTACCCAAC</u>
HUMAN		CAGACCAGCAACTTCCAAAATCTGCCCCGTG--TTTAGCCTTGTTAAAGGGGTGTCTCA	
MOUSE		<u>CGGACCAACAAC</u>	<u>TCCAGAAGGTTCTGCGAGGCATAGAGCCATTCCGTAGGGACATCTCG</u>
HUMAN		CTCCTTCAGGGAAGTGGAAAAG--GGGATCTGATTA----TTGAGGTGTGAAGGAAT	Forkhead-like
MOUSE		GTGCTTCTGAGGAAGCGGACCGAGCAGGGATCCGATGACGACTGGAGATGTTGAAGGAAT	<u>GTGCTTCTGAGGAAGCGGACCGAGCAGGGATCCGATGACGACTGGAGATGTTGAAGGAAT</u>
HUMAN		AAATAATCAGTCCACAAATAAACAAACTGTCC--GGGATTCTCTAGAGGGAAGGAGAAATC	Forkhead-like
MOUSE		<u>AAATA-CCAGTCCACAAATAAACAAACTGTCCCCGGGATTCTCTAGAGGGAAGGAGCACGC</u>	Ets-like
HUMAN		CTTGAAG-----GAGATCCAAGTCGCTCCAGGTCTGCCCTGCCCGAATAATATC	
MOUSE		TTGAAGGTCGGGGAACCTCCGAGTCGCTGTGCGTCAAGGT-TGGCATAAAATT	

475

